The Influence of Randomized Controlled Trials on Development Economics Research and on Development Policy

Paper prepared for "The State of Economics, The State of the World"
Conference proceedings volume

Abhijit Vinayak Banerjee
Esther Duflo
Michael Kremer[12]

September 11, 2016

Many (though by no means all) of the questions that development economists and policymakers ask themselves are causal in nature: What would be the impact of adding computers in classrooms? What is the price elasticity of demand for preventive health products? Would increasing interest rates lead to an increase in default rates? Decades ago, the statistician Fisher proposed a method to answer such causal questions: Randomized Controlled Trials (RCT) (Fisher, 1925). In an RCT, the assignment of different units to different treatment groups is chosen randomly. This insures that no unobservable characteristics of the units is reflected in the assignment, and hence that any difference between treatment and control units reflects the impact of the treatment. While the idea is simple, the implementation in the field can be more involved, and it took some time before randomization was considered to be a practical tool for answering questions in social science research in general, and in development economics more specifically.

About twenty years ago, the idea of randomized controlled trials was just starting to make its way into development economics. Starting in 1994, Glewwe, Kremer, and Moulin (2009) kick-started the use of randomized evaluations among develop xent economists and practitioners (Kremer, 2003). In 1997, the PROGRESA randomized controlled trial began, marking the first evaluation of a large scale policy effort in a developing country. With the launch of these randomized evaluations, we, perhaps naively, expressed the hope that randomized controlled trials would revolutionize social policy in the twenty first century, much as they had revolutionized medicine in the 20th century (Duflo and Kremer, 2003; Duflo, 2004; Banerjee et al, 2007). With the century less than 20 years old, it seems a little premature to evaluate this claim. Randomized evaluations clearly take a larger place in the policy conversation now than they did at the turn of the century and receive substantially more funding from donor organizations and local governments. Policy innovations that have been tested with RCTs have reached millions of people. However, the amount of money involved is still small. Development policy, moreover, is known for its twists and turns; many have predicted that RCTs are just the current fad and, soon enough, will have their comeuppance.

Something that we did not anticipate, however, has undoubtedly happened: randomized controlled trials have, if not revolutionized, at least profoundly altered, the practice of development economics as an academic discipline. Some scholars applaud this change (we are obviously in that camp), while others rue it (Ravallion, 2012; Deaton, 2010) but the fact is not really in dispute. In this essay, we start by quantitatively documenting this remarkable evolution. We go on to discuss the ways in which the field has been affected by the practice of randomized controlled trials, and what we see as their main contributions to the practice of development economics.

The popularity of RCTs as a research tool has sometimes been seen as conflicting with their potential (or ambition) for changing the world. The view is that the "academic" desire to come up with the cleverest research design may not line up with the practitioners need to identify scalable innovations (the next cell phone), or change "systems" (health care) or reform institutions (democracy). Using the USAID Development Innovation Ventures (DIV) portfolio as a case study, we identify the policy innovations tested with DIV funding that have eventually led to large scale reach (over 100,000 people). The analysis suggests that the proposed opposition between interesting and important is not particularly pertinent. In practice, many of the interventions supported by DIV that have reached this scale started as small research projects driven by academics. These projects also had the greatest "bang for the buck" evaluated in terms of life eventually reached per USAID initial funding dollars.[3] We conclude this essay by discussing what this tells us about the policy process and the role RCTs can have in it.

---

[3] Which does not necessarily imply they have the highest social return.

1. Rapid growth

Over the last fifteen years, the use of experiments has expanded in academia and in international organizations: the DIME group at the World Bank lists over 200 studies, the vast majority of them randomized, and Arianna Legovini, the head of DIME, estimates that if we take the Bank as a whole, there are at least 475 randomized controlled studies going on (Legovini, personal communication). Tables 1 and 2 and Figures 1 to 5 summarize some of the trends in the use of experiments over time.

We start by a review of impact evaluations conducted by Cameron et. al. (2015) (Figures 1 and 2). They compiled a repository of 2259 impact evaluation studies in development economics that were published between 1981 and 2012 by searching all major academic databases in health, economics, public policy and the social sciences. They supplemented this with an online crowdsourcing effort, which offered a $10 gift certificate per qualifying paper that was not already in the database. They then classified them by sector and by type. Overall, 66% (1491) of those evaluations are RCTs. Figure 2 shows that the number of RCTs has grown rapidly over time.

Next, we look at the data compiled by Aidgrade (Vivalt, 2015). Aidgrade compiles the results of impact evaluations of development interventions. According to Vivalt:

> The evaluations included in the AidGrade database were carefully selected from a number of different databases and online sources, the detailed process for which is outlined in Vivalt (2015). AidGrade.org employees first chose 30 topics they felt were important development issues. Those lists were combined and made into one large list of topics. The list was then narrowed down based on whether or not there were likely to be enough evaluations for a meta-analysis. The search universe includes search aggregators, such as Google Scholar and EBSCO, but also includes the J-PAL, IPA, CEGA, and 3ie only databases.

Figure 3 shows the number of evaluations per year, and Figure 4 shows how the evaluations ventilate over time between RCTs in economics, RCTs in other fields (e.g. medical trials) and non-RCTs. Both figures show a clear trend in both the number and the fraction of RCTs among the impact evaluations that are surveyed. RCTs are particularly popular among younger researchers. Figures 4 and 5 show the number and the fraction of researchers who carry out RCT among the fellows and associates of BREAD, the association of development economists, by year in which they have obtained their PhD. The number clearly increases among the recent PhDs, and while this is in part driven by a larger number of recent fellows and associates, the fraction of them who do RCTs increases as well.

The number of RCTs presented at development economics conferences grew rapidly until 2010, and stabilized (or decreased) after that. At the annual conference of BREAD (the flagship conference in development economics), the fraction of papers featuring RCTs increased from 8% in 2005 to 63% in 2010, and hovered around 40-50% after that (except for the last conference, at Georgetown, where it was 28%). At NEUDC, a larger conference attended by many junior researchers, the fraction of RCTs has been fairly stable, ranging between 16% and 24% for the years 2012 to 2016 (the years for which we could get the papers) and showing no particular trend (Table 1).

RCTs have made a clear entry in top academic journals. Looking at AER, QJE, Econometrica, Review of Economic Studies and JPE, the number of RCT studies was 0 in 1990, 0 in 2000, and 10 in 2015 (Table 2). At the same time, the number of development papers published in these journals almost doubled (from 17 in 1990 to 32 in 2015). Table 2 also provides the details by journal. This is not driven by any particular journal (except that *Econometrica* does not seem to contribute much). Note that this does not mean that RCT studies have supplanted other type of work: the vast majority of published work on development is still non-RCT (if we take lower ranked journals), and even in top journals the experiments have been in addition to the (limited number of) papers that were published on development.

Beyond the growth in the number of experiments, and in the number of researchers who carry them out, what also stands out is the range and the ambition of the projects that are attempted: few topics seem off limit, and scale does not seem to be a barrier.

Researchers work directly with governments to randomize aspects of their work. Finnan, Olken and Pande (2015) describe several of these ambitious experiments. For example, Dal Bó, Rossi and Finan (2013) randomize the wages at which new government employees are hired, while Olken, Khan and Khwaja (2016) randomize incentives for tax collectors in Pakistan and Ashraf, Bandiera, & Lee (2015) work on how government health workers are recruited for their job. In experiments covering several districts and millions of workers Muralidharan, Niehaus and Sukhtankar (2016) and Banerjee et al. (2016) evaluate two separate process changes in the payment of wages of India's major workfare program (MGNREGS), while Banerjee, Duflo and Keniston (2014) randomize reforms in the police department in India and Duflo, Greenstone, Pande, and Ryan (2013a and 2013b) randomize the enforcement of pollution regulation on industrial firms in India.

Researchers work at a scale that is sufficient to capture market equilibrium effects: Muralidharan and Sundararaman (2015) randomize a private school voucher at the school market level while Muralidharan, Niehaus and Sukhtankar (2016), in their aforementioned experiment, are able to look at the impact of MGNREGS on wages and productivity.

The range of topics keeps expanding. Development economists study alcohol addiction (Schilbach, 2015), electoral fraud in Afghanistan (Callen & Long, 2014), Cognitive Behavioral Therapy (CBT) for ex-combatants (Blattman et al., 2015), early childhood stimulation and development (Attanasio et al., 2014).

In summary, randomized experiments have become, not so much the "gold standard" as just a standard tool in the toolbox. Running an experiment is now sufficiently commonplace that by itself it does not guarantee that the paper would get into a top journal or even the BREAD conference. On the other hand, researchers from all sorts of perspectives have come to consider RCTs as a feasible option for answering the questions they are interested in. This level of comfort is in part due to the growth of several entities that help researchers with the fieldwork including by codifying and standardizing experimental practices, training enumerators, etc. The leader for this is Innovation for Poverty Action (IPA), with its vast network of country offices and experienced staff workers, but also J-PAL, CEGA, and the World Bank. There is also more funding available, from USAID (DIV in particular), the World Bank (SIEF and DIME), DFID, The Bill and Melinda Gates Foundation, The William and Flora Hewlett Foundation, The International Initiative for Impact Evaluation, in particular and, more recently, the Global Innovation Fund. But part of it also has to do with the appeal of the technique. In the next section, we reflect on the influence that RCTs have had on development economics research and why.

2. The influence of RCTs on development economics research

The remarkable growth in the number of RCTs, and more generally in the importance of empirical development economics as a field, is in itself a dramatic change. The type of development research that is carried out today is significantly different from research conducted even fifteen years ago. A reflection of this fact is that many researchers who were openly skeptical of RCTs, or simply belonged to an entirely different tradition within development economics (e.g. Daron Acemoglu, Derek Neal, Martin Ravallion, Mark Rosenzweig) are involved in one or more randomized controlled trials in a developing country.

Early discussions of the merits (or lack thereof) of randomization put a lot of emphasis on its role in the reliable identification of internally valid causal effects and the external validity of such estimates. We, and others, have had these discussions in other places (Heckman 1992; Banerjee, 2007; Duflo, Glennester, and Kremer, 2007; Banerjee and Duflo, 2009; Deaton, 2010), and we won't reproduce them here. As we began to argue in Banerjee and Duflo (2009), we actually think that these discussions somewhat miss the point about why RCTs are really valuable, and why they have become so popular with researchers.

a.   A greater focus on identification across the board

The original motivation of randomized experiments, starting with Neyman (1923) (as a theoretical device) and Fisher (1925) (who was the first to propose physically randomizing units), was a focus on the credible identification of causal effects. As Athey and Imbens (2016) write in their chapter for *The Handbook on Field Experiments*:

> There is a long tradition viewing randomized experiments as the most credible of designs to obtain causal inferences. Freedman (2006) writes succinctly 'experiments offer more reliable evidence on causation than observational studies.' On the other hand, some researchers continue to be skeptical about the relative merits of randomized experiments. For example, Deaton (2010) argues that 'evidence from randomized experiments has no special priority. …Randomized experiments cannot automatically trump other evidence, they do not occupy any special place in some hierarchy of evidence.' Our views align with that of Freedman and others who view randomized experiments as playing a special role in causal inference. Whenever possible, a randomized experiment is unique in the control that the researcher has over the assignment mechanism, and by virtue of this control, selection bias in comparisons between treated and control units can be eliminated. That does not mean that randomized experiments can answer all causal questions. There are a number of reasons why randomized experiments may not be suitable to answer particular questions,

For a long time, observational studies and randomized studies progressed on largely parallel paths: in agricultural science, and then biomedical studies, randomized experiments were quickly accepted, and a vocabulary and statistical apparatus to think about them were developed. Despite the adoption of randomized studies in other fields, in the social sciences most researchers continued to reason exclusively in terms of observational data. The main approach was to estimate associations, and then to try to assess the extent to which these associations reflect causality (or to explicitly give up on causality). Starting with Rubin's (1974) fundamental contribution, researchers started to use the experimental analog to reason about observational data, and this set the stage for thinking about how to analyze observational data through the lens of the "ideal experiment".

Through the 1980s and 1990s, motivated by this clear thinking about causal effects, labor economics and public finance were transformed by the introduction of new empirical methods for estimating causal effects (matching, instrumental variables, difference-in-differences and regression discontinuity designs). Development economics also embraced those methods starting in the 1990s, but unlike in labor economics and public finance, some researchers also decided that it may be possible to go straight to the "ideal" experiment or to go back and forth between experimental and non-experimental studies. This means that the two literatures developed in close relationship, constantly cross-fertilizing each other.

The non-experimental literature was completely transformed by the existence of this large RCT movement. When the "gold standard" is not just a twinkle in someone's eyes, but the clear alternative to a particular empirical strategy and a benchmark for it, researchers feel compelled to think harder about identification strategies, and to be more inventive and rigorous about them. As a result, researchers have

become increasingly more clever at identifying and using natural experiments, and at the same time, much more cautious in interpreting the results from them. Not surprisingly therefore, the standards of the non-experimental literature have improved tremendously over the last few decades, without necessarily sacrificing their ability to ask broad and important questions. To take some examples, Alesina, Guliano and Nunn (2013) use suitability to the plow to study the long run determinants of the social attitudes towards the role of women; Padro-i-Miguel, Qian and Yao (2014) use a difference and difference strategy to study village democracy; and Banerjee and Iyer (2005) and Dell (2010) use a spatial discontinuity to look at the long run impact of extractive institutions. In each of these cases, the questions are approached with the same eye for careful identification as other more standard program evaluation questions.

Meanwhile, the RCT literature was also influenced by work done in the non-experimental literature. The understanding of the power (and limits) of instrumental variables allowed researchers to move away from the basic experimental paradigm of the completely randomized experiment with perfect follow up and use more complicated strategies, including encouragement designs. Techniques developed in the non-experimental literature offered ways to handle situations in the field that are removed from the ideal setting of experiments (imperfect randomization, non-compliance, attrition, spillovers and contamination, etc.). Structural methods were combined with experiments to estimate counterfactual policies (Todd and Wolpin, 2006; Attanasio, Meghir and Santiago, 2011).

More recently, machine learning techniques have also been combined with experiments to model treatment effect heterogeneity (see Athey and Imbens (2016) for a recent review of the econometrics of experiments).

Of course, the broadening offered by these new techniques comes with the cost of making additional assumptions on top of the original experimental assignment, and those assumptions may or may not be valid. This means that the difference in the quality of identification between a very well-identified, non-experimental study and a randomized evaluation that ends up facing lots of constraints in the field or tries to estimate parameters beyond pure treatment effects is a matter of degree. In this sense, there has been a convergence across the empirical spectrum in terms of the quality of identification, though mostly because experiments have pulled the remaining study designs up with them.

b. Assessing external validity

In the words of Athey and Imbens (2016): "external validity is concerned with generalizing causal inferences, drawn for a particular population and setting, to others, where these alternative settings could involve different populations, different outcomes, or different contexts." The question of the external validity of RCTs is even more hotly debated than that of their internal validity. This is perhaps because, unlike internal validity, there is no clear endpoint to the debate: heterogeneity in treatment effects across different types of individuals could always occur, or heterogeneity in the effect may result from ever-so-slightly different treatments. As Banerjee, Chassang and Snowberg (2016) acknowledge: "External policy advice is unavoidably subjective. This does not mean that it needs to be uninformed by experimental evidence, rather, judgment will unavoidably color it."

It is worth noting that there is very little here that is specific about RCTs (Banerjee, 2009). The same problem afflicts all empirical analysis with the one exception of what Heckman (1992) calls the "randomization bias." "Randomization bias" refers to the fact that experiments require the consent of both the subjects and the organization who is carrying out the program, and these people may be quite different. Glennerster (2016), in her chapter in the *Handbook of Field Experiments*, provides the list of the characteristics of the ideal partner, and they are clearly not representative of the typical NGO. On the other hand, it is worth pointing out that any naturally occurring policy that gets evaluated (i.e. not an RCT) is also selected: the evaluation requires that the policy did take place, and that was presumably because someone thought it was a good idea to try it out.

In general, any study takes place in a particular time and place, and that would affect results. This does not imply that subjective recommendations by experts, based both on their priors and the results of their experiments, should not be of some use for policymakers. Most policymakers are not stupid, and they do know how to combine the data that is presented to them with their own prior knowledge of their settings. From our experience, we have often observed that when presented with evidence from an RCT on a program of interest, the immediate reaction of a policymaker is to ask whether an RCT could be done in their own context.

There is one clear advantage that RCTs do offer for external validity, although it is not often discussed, and has not been systematically exploited as yet. To assess any external validity issues, it is helpful to have well-identified causal studies in multiple settings. These settings should vary in terms of the distribution of characteristics of the units, and possibly in terms of the specific nature of the treatments or the treatment rate, in order to assess the credibility of generalizing to other settings. With RCTs, because we can, in principle, control where and over what sample experiments take place (and not just how to allocate the treatment within a sample), we can therefore, get a handle on how treatment effects might vary by context. Of course, on its own, this is not sufficient to say anything much, if we account for infinite unstructured variation in the world. But there are several ways to make progress.

A first approach is to combine existing evaluations, and make assumptions about the possible distribution of treatment effects. Rubin (1981) proposes modeling treatment effect heterogeneity as stemming from a normal distribution: in each site, the causal effect of the treatment is a site-specific effect drawn from a normal distribution. The goal is to estimate the mean and the variance of the treatment effect, and the implied specific site effect, taking into account the fact that we have the other effects too. An interesting case study is the effect of microfinance programs. Meager (2016) analyzes data from seven randomized experiments, including six published in a special issue of the *American Economic Journal: Applied Economics* in 2015, and finds remarkable consistency in the mean effects across these studies, but much more heterogeneity in their variance. Of course, to carry out this exercise properly, we need access to an un-selected sample of studies, and since there is publication bias in economics, the sample of published studies may not be representative of all the studies that exist. This is where another advantage of RCT kicks in: since they have a defined beginning and end, they can in principle be registered. To this end, the American Economic Association recently created a registry of randomized trials (www.socialscienceregistry.org), which, as of June 1, listed 699 studies. The hope is that all projects are registered, preferably before they are launched, and that results are clearly linked to the study, so that in the future meta-analysts can work from the full universe of studies.

A second approach is to conceive projects as multi-site projects from the start. One recent example of such an enterprise is the "Graduation" approach—an integrated, multi-faceted program with livelihood promotion at its core that aims to "graduate" individuals out of extreme poverty and onto a long-term, sustainable higher consumption path. BRAC, the world's largest nongovernmental organization, has scaled-up this program in Bangladesh (Bandiera et al. 2013), while NGOs around the world have engaged in similar livelihood-based efforts. Six randomized trials were undertaken over the same time period across the world (Ethiopia, Ghana, Honduras, India, Pakistan, and Peru). The teams regularly communicated with each other and with BRAC to ensure that their local adaptations remain true to the original program. The results suggest that the integrated multi-faceted program was "sufficient" to increase long-term income, where long-term is defined as three years after the productive asset transfer (Banerjee et al., 2015). Using an index approach to account for multiple hypotheses testing, positive impacts were found for consumption, income and revenue, asset wealth, food security, financial inclusion, physical health, mental health, labor supply, political involvement and women's decision-making after two years. After a third year, the results remained the same in 8 out of 10 outcome categories. There is country-by-country variation (e.g. the program was ineffective in Honduras), and the team is currently working on a meta-analysis to quantify the level of heterogeneity.

One issue is that there is little that the researcher can do ex-post to reliably identify the source of differences in findings across countries. A third possible approach would be to take guidance from the first few sites to make a prediction on what the next sites would find. To discipline this process, researchers would be encouraged to use the results from existing trials to make some explicit predictions about what they expect to observe in other samples (or with slightly different treatments). These can serve as a guide for subsequent trials. This idea is discussed in Banerjee, Chassang and Snowberg (2016), who call it "structured speculation." They propose the following broad guidelines for structured speculation:

1. Experimenters should systematically speculate about the external validity of their findings.
2. Such speculation should be clearly and cleanly separated from the rest of the paper, maybe in a section called "speculation"
3. Speculation should be precise, and falsifiable

Structured speculation has three advantages, according to Banerjee, Chassang and Snowberg (2016). First, it ensures that the researcher's specific knowledge is captured. Second, it creates a clear sense of where else experiments should be run. Third, it creates incentives to design research that has greater externality. They write:

> To address scalability, experimenters may structure local pilot studies for easy comparison with their main experiments. To identify the right sub-populations for generalizing to other environments, experimenters can identify ahead of time the characteristics of groups that can be generalized, and stratify on those. To extend the results to populations with a different distribution of unobserved characteristics, experimenters may elicit the former using the selective trial techniques discussed in Chassang et al. (2012), and run the experiments separately for each of the groups so identified.

As this approach was just proposed recently, there are few examples as yet. A notable example is Dupas (2014). Dupas (2014) studies the effect of short-term subsidies on long-run adoption of new health products, and reports that short-term subsidies had a significant impact on the adoption of a more effective and comfortable class of bed nets. The paper then provides a clear discussion of external validity: It first spells out a simple and transparent argument relating the effectiveness of short-run subsidies to: 1) the speed at which various forms of uncertainty are resolved; 2) the timing of user's costs and benefits. If the uncertainty over benefits is resolved quickly, short-run subsidies can have a long-term effect. If uncertainty over benefits is resolved slowly, and adoption costs are incurred early on, short-run subsidies are unlikely to have a long-term effect.

It then answers the question "For what types of health products and contexts would we expect the same results?" It does so by classifying potential technologies into three categories based on how short-run (or one-time) subsidies would change adoption patterns. Clearly, there could be such discussions at the end of all papers, not just ones featuring RCTs. But because RCTs can be purposefully designed and placed, there is a higher chance of follow-up in this case.

c. Observing the unobservable

If the main benefit of randomization is not the identification of causal effect, what is it? And what explains its remarkable success among researchers?

We agree with Athey and Imbens (2016) that "a randomized experiment is unique in the control that the researcher has over the assignment mechanism" and we would take the argument one step further: randomization is also unique in the control that the researcher (often) has on the treatment itself. In observational studies, however beautifully designed, the researcher is limited to evaluating what has been implemented in the world. In a randomized experiment, she can manipulate the treatment in ways that we do not observe in reality. This has a number of advantages. First, she can innovate, i.e. design new policies or interventions that she thinks will be effective based on prior knowledge or theory, and test them even if no policymaker is thinking of putting them in practice yet. Development economists have many ideas, often inspired by what they have read or researched, and many of the randomized experiment projects come out of those: they test in the field an intervention that simply did not exist before (a kilogram of lentil for parents who vaccinate their kids; stickers to encourage riders to speak up against a bad driver; free chlorine dispensers, etc.).

Second, she can introduce variations that will help her establish facts that could not otherwise be established. The well-known Negative Income Tax (NIT) Experiment was designed with precisely that idea in mind: in general, when wages are raised, this creates both income and substitution effects which cannot easily be separated (Heckman, 1992). But randomized manipulation of the slope and the intercept of a wage schedule makes it possible to estimate both together. Interestingly, after the initial NIT and the Rand Health Insurance Experiment, the tradition of social experiments in the US, as Judy Gueron describes in her chapter in the *Handbook of Field Experiments* (Gueron, 2016), has mainly been to obtain causal effect of social policies that were often fairly comprehensive packages. In contrast, development economists have worked both on evaluations of real policies (e.g. the PROGRESA evaluation, or, more recently, the evaluation of the graduation program) but also on what Congdon, Kling, Ludwig, and Mullainathan (2016) describe as "mechanism experiments." They write:

> Broadly, a mechanism experiment is an experiment that tests a mechanism—that is, it tests not the effects of variation in policy parameters themselves, directly, but the effects of variation in an intermediate link in the causal chain that connects (or is hypothesized to connect) a policy to an outcome. That is, where there is a specified policy that has candidate mechanisms that affect an outcome of policy concern, the mechanism experiment tests one or more of those mechanisms. There can be one or more mechanisms that link the policy to the outcome, which could operate in parallel (for example when there are multiple potential mediating channels through which a policy could change outcomes) or sequentially (if for example some mechanisms affect take-up or implementation fidelity). The central idea is that the mechanism experiment is intended to be informative about some policy but does not involve a test of that policy directly.

In other words, mechanism experiments do not confine themselves to testing feasible (or desirable) policies. For example, cars with broken windows could be put in the street to test the broken window theory. Once we realize that we are not limited to a set of realistic policy options (though we are constrained by what is ethically acceptable), this opens up a wide range of possibilities.

Banerjee and Duflo (2009) discuss some examples of mechanism experiments. One prominent example in development is a project conducted by Karlan and Zinman (2005) in collaboration with a South African lender that gives small loans to high-risk borrowers at high interest rates. The experiment was designed to test the relative weights of *ex post* repayment burden (including moral hazard) and *ex ante* adverse selection in loan default. Potential borrowers with the same observable risk are randomly offered a high or a low interest rate in an initial letter. Individuals then decide whether to borrow at the solicitation's offer rate. Of those who apply at the higher rate, half are randomly offered a new, lower contract interest rate when they are actually given the loan, whereas the remaining half continue at the offer rate. Individuals did not know *ex ante* that the contract rate could differ from the offer rate. The researchers then compared repayment performance of the loans in all three groups. The comparison of those who responded to the high-offer interest rate with those who responded to the low offer interest rate in the population that received the same low contract rate allows the identification of the adverse selection effect; comparing those who faced the same offer rate but differing contract rates identifies the repayment burden effect. The basic idea of varying prices *ex post* and *ex ante* to identify different parameters has since been replicated in several different studies (e.g. Ashraf, Berry and Shapiro (2010) and Cohen and Dupas (2010)). The experimental variation was key here, and not only to avoid bias: In the world, we are unlikely to observe a large number of people who face different offer prices, but receive the same actual price.

Experiments can also be set up to understand the way institutions function. An example is Bertrand et al. (2007), who set up an experiment to understand the structure of corruption in the process of obtaining a driving license in Delhi. They recruited people who are aiming to get a driving license and set up three groups, one that receives a bonus for obtaining a driving license quickly, one that gets free driving lessons, and a control group. They found that those in the "bonus" group get their licenses faster, but those who get the free driving lessons do not. They also found that those in the bonus group are more likely to pay an agent to get the license (who, they conjecture, bribes someone). They also found that the applicants who hired an agent were less likely to have taken a driving test before getting a driving license. Although they did not appear to find that those in the bonus group who get licenses are systematically less likely to know how to drive than those in the control group (which would be the litmus test that corruption does result in an inefficient allocation of driving licenses), this experiment provides suggestive evidence that corruption in this case does more than "grease the wheels" of the system.

Such designs do not always directly lead to actionable policy, but they have allowed us to describe or understand how the world works. For example, in the seminal Bertrand and Mullainathan (2004) study, researchers sent resumes to prospective employers. The resumes are paired, such that there are identical resumes, except for the name of the job applicants, who can either be white sounding or African American sounding. They find that "applicants" with black sounding names are half as likely to be called back as those with white sounding names. Furthermore, being highly educated does not help, which

suggests that something other than statistical discrimination is at play. This design has been replicated hundreds of times in different settings, providing extensive evidence of discrimination against different people and in different markets. This large body of evidence does not necessary point to a specific solution to this problem, or even helps us determine the root of this behavior, but, unlike the previous literature, it provides clear evidence that the phenomenon exists.

d. Data collection

Experiments have also spurred creativity in measurement. In principle, there is no automatic link between careful and innovative collection of microeconomic data and the experimental method. And, indeed, there is a long tradition in development economics to collect data that is specifically designed to test theories: both the breadth and the quantity of microeconomic data collected in development economics has exploded in recent decades, not only in the context of experiments (see Udry (1995) for a prominent early example).

However, one specific feature of experiments that serves to encourage the development of new measurement methods is high take-up rates and a specific measurement problem. In many experimental studies, a large fraction of those who are intended to be affected by the program are actually affected. This means that the number of units on which data needs to be collected to assess the impact of the program does not have to be very large and that data are typically collected especially for the purpose of the experiment. Elaborate and expensive measurement of outcomes is therefore easier to obtain than in the context of a large multipurpose household or firm survey. By contrast, observational studies must often rely on variation for identification (policy changes, market-induced variation, natural variation, supply shocks, etc.) that cover large populations, requiring the use of a large data set often not collected for a specific purpose. This makes it more difficult to fine-tune the measurement to the specific question at hand. Moreover, even if it is possible *ex post* to do a sophisticated data collection exercise specifically targeted to the question, it is generally impossible to do it for the preprogram situation. This precludes the use of a difference-in-differences strategy for these types of outcomes, which again limits the incentives to collect them ex-post.

Some of the most exciting recent developments in empirical development economics have to do with measurement. Researchers have turned to other sub-fields of economics, as well as entirely different fields, to borrow tools for measuring outcomes. Examples include soil testing and remote sensing in agriculture (see de Janvry, Sadoulet and Suri (2016) for a review on agriculture); techniques developed by social psychologists for difficult to measure outcomes, such as audit and correspondence studies, implicit association tests, Goldberg experiments and List experiments (see Bertrand and Duflo (2016) for a review of their use to measure discrimination); tools developed by cognitive psychologists for child development (Attanasio et al., 2014); tools inspired by economic theory, such as Becker-DeGroot-Marshak games to infer willingness to pay (see a discussion in Dupas and Miguel, 2016); biomarkers in health, beyond the traditional height, weight and hemoglobin (cortisol to measure stress for example); wearable devices to measure mobility or effort (Rao, Schilbach & Schofield, in progress; Kreindler, in progress).

Specific methods and devices that exactly suit the purpose at hand have also been developed for experiments. Olken (2007) is one example of the kind of data that can be collected in an experimental setting. The objective was to determine whether audits or community monitoring were effective ways to curb corruption in decentralized construction projects. Getting a reliable measure of actual levels of corruption was thus necessary. Olken focused on roads and had engineers dig holes in the road to measure the material used. He then compared that with the level of material reported to be used. The difference is a measure of how much of the material was stolen, or never purchased but invoiced, and thus an objective measure of corruption. Olken then demonstrated that this measure of "missing inputs" is affected by the threat of audits, but not, except in some circumstances, by encouraging greater attendance at community meetings. Rigol, Hussam & Regianni (in progress) provide another example of clever data collection methods. For their experiment, they designed soap dispensers that could track when the pump was being pushed in order to accurately measure if and when people wash their hands, and hired a Chinese company to manufacture the dispensers. Similar "audit" methodologies are used to measure the impact of interventions in health, such as patients posing with specific diseases to measure the impact of training (Banerjee et al., 2016) or ineligible people attempting to buy free bed nets (Dupas et al., 2016). Even a partial list of such examples would be very long.

In parallel, greater use is being made of administrative data, which are often combined with large-scale experiments. For example, Banerjee et al., (2016) make use of both publicly available administrative data on a workfare program in India and restricted expenditure data made available to them as part of the experiment; Khan, Khwaja and Olken (2016) use administrative tax data; and Attanasio, Medina and Meghir (2016) use unemployment insurance data to measure the long term effect of job training in Colombia.

The bottom line is that there has been great progress in our understanding of how to creatively and accurately collect or use existing data that go beyond the traditional survey, and these insights have led both to better projects and to innovations in data collection that have been adopted in non-randomized work as well.

e.   Iterate and build on previous research in the same settings

The next methodological advantage of RCTs also relates to the control that researchers have over the assignment and, often enough, over the treatments themselves. Well-identified policy evaluations often leave us with many questions on why things turned out the way they did. For example, a number of papers using regression discontinuity designs find that the impact of "elite" schools on the marginal child who is admitted tends to be very low. These results seem to hold both in rich and in poor countries (Abdulkadiroglu , Angrist, and Pathak, 2014; Dobbie and Fryer, 2014; Lucas and Mbiti, 2014; Dustan, de Janvry, and Sadoulet, 2015; Clark, 2009). But this leaves a number of questions pending: does this mean that the impact is zero for all students or just the marginal student? Is it because peers don't matter and curriculum doesn't matter or because they both matter but cancel out?

While some progress can be made (for example, Abdulkadiroglu, Angrist & Pathak (2014) exploit the fact that students take two different tests to get a handle on the impact of the program for different types of students), one is necessarily limited by the type of policy variation that is actually available. The result from a single RCT often likewise raises more questions than it can actually answer. For example, when Duflo, Kremer and Robinson (2008a) found that the return to fertilizer appears to be very large even when used by the farmers themselves on their own fields (and not just on experimental plots), one possible policy response might have been to follow Jeff Sachs' idea in distributing fertilizer for free. But this was not their next step: instead they started wondering why farmers are not using more fertilizer. This set them down a path that led them to set up a number of experiments in the same setting: some focused on learning and social networks, and some on the difficulty to save even over short periods of time. This latter inquiry led them down the path of designing and implementing a specific product where the household was offered the option of buying fertilizer in advance (Duflo, Kremer and Robinson, 2008b). The social network interventions found surprisingly little diffusion of agricultural innovation to immediate friends, and this set them down another path: how could it be the case, given all we know about how much people talk about agriculture? To unpack this further, they introduced a simple device designed to address a problem that they noticed in their first set of experiments: households tend to overuse fertilizer (conditional on using it), relative to what appears to be profit-maximizing. They then set up a number of experiments to study in what conditions this device does spread, and what this tells us about how farmers decide whether to talk to each other and trust each other (Duflo, Kremer, Robinson, Schilbach, in progress).

Analyzing these results will no doubt spur new questions and experiments. All empirical science is of course iterative, with studies building on each other. But the ability to work in the same setting, with the same outcome and measurement, is extremely precious, and not available outside of a controlled setting.

f. Unpacking the interventions

Finally, RCTs, allow the possibility to "unpack" a program, to its constituent elements. Here again the work may be iterative. For example, all the initial evaluations of the BRAC ultra poor program were done using their "full package," as were a large number of evaluations of the Mexican CCT program PROGRESA. But both for research and for policy, once we know that the full program works, there is a clear interest in knowing why it works. In recent years, a number of papers have looked "inside" CCT, relaxing the conditionality, for example. Some work has been conducted on the role and the type of conditionality, (see Baird, McIntosh and Ozler (2011), Bursztyn and Coffman (2012) and Benhassine et al. (2015) for examples), followed by many papers experimentally varying other features (we return to the impact of this work below).

Similarly, the early results of the evaluation of the ultra poor program have set the stage, both for a more theoretically grounded understanding of exactly which market failures led to a poverty trap, as well as a more practically grounded understanding of whether all of the interventions were truly necessary or if certain components could be removed. In the event that some components are unnecessary, costs could be lowered considerably, allowing the program to reach more people using the same budget. Hanna and Karlan (2016) discuss how one could go from the initial "full package" evaluation to this greater understanding:

> The ideal method, if unconstrained by budget and organizational constraints, is a complex experimental design that randomizes all permutations of each component. The productive asset transfer, if the only issue were a credit market failure, may have been sufficient to generate these results, and if no other component enabled an individual to accumulate sufficient capital to acquire the asset, the transfer alone may have been a necessary component. The savings component on the other hand may have been a substitute for the productive asset transfer, by lowering transaction costs to save and serving as a behavioral intervention which facilitated staying on task to accumulate savings. Clearly it is not realistic in one setting to test the necessity or sufficiency of each component, and interaction across components: Even if treated simplistically with each component either present or not, this would imply 2x2x2x2 = 16 experimental groups.

Several studies have tackled pieces of the puzzle, and more are underway (see the review in Hanna and Karlan, 2016). The way forward is clearly going to be the development of a mosaic, rather than any one definitive study that both tests each component and also includes sufficient contextual and market variations that it can help set policy for a myriad of countries and populations. More work is needed to tease apart the different components: asset transfer (addresses capital market failures), savings account (lowers savings transaction fee), information (addresses information failures), life-coaching (addresses behavioral constraints, and perhaps changes expectations and beliefs about possible return on investment), health services and information (addresses health market failures), consumption support (addresses nutrition-based poverty traps), etc. Furthermore, for several of these questions, there are key, open issues for *how* to address them; for example, life-coaching can take on an infinite number of manifestations. Some organizations conduct life-coaching through religion, others through interactive

problem-solving, and others through psychotherapy approaches (Bolton et al., 2003; Bolton et al., 2007; Patel et al., 2010). Much remains to be learned not just about the promise of such life-coaching components, but how to make them work (if they work at all).

In some settings, particularly when working on a large-scale with a government, it is actually possible to experiment from the beginning with various versions of a program. This serves two purposes: it gives us a handle on the theory behind the program, and it has operational value for the government, who can pick the most cost-effective combination. Banerjee et al. (2015) is an example of this approach. The government of Indonesia was interested in reducing corruption in their rice distribution program (Raskin), which is infamous for reaching few of its intended beneficiaries and for not always been sold at the right price. They thought that delivering a card to the beneficiaries with the eligibility information might ameliorate this problem, and lead to greater benefits. Working with the Government of Indonesia, the authors designed a set of field experiments to provide information directly to eligible households. In 378 villages (randomly selected from among 572 villages spread over three provinces), the central government mailed "Raskin identification cards" to eligible households to inform them of their eligibility and the quantity of rice that they were entitled to. To unbundle the mechanisms through which different forms of information may affect program outcomes, the government also experimentally varied how the card program was run along three key dimensions— whether an additional rule (the copay price) was also listed on the card, whether information about the beneficiaries was also made very public, and whether cards were sent to all eligible households or only to a subset. The researchers then collected data on eligible and ineligible rice purchases and price paid for all villages. On net, they found that the card did lead to large increases in the amount of subsidies received by the households. Further, they found that the information on the card matters: the price paid was lower when the price was indicated on the card. They also found that the card was more effective when the information was made public. Finally, public information is not sufficient on its own, the physical card also matters.

Knowing all of this is important to understand the mechanisms at play. It was also immediately actionable for the government, who proceeded to scale-up the program, and to provide cards with price information, to all eligible households accompanied by posters. Cards were distributed to over 65 million individuals. This is one occasion where the researchers' and the government's interests were exactly aligned. Is it more generally true?

3. Have RCTs become too academic to lead to any real world changes?

RCTs have changed development economics but have they also had significant influence in the world? If RCTs are pushing forward the frontiers of academic research by seeking to understand mechanisms and testing ideas generated by academics themselves, does this make them too academic and less useful for policy?

In this section we argue that RCTs can contribute to policy not only by providing evidence on specific programs that can be scaled, but also by changing the general climate of thinking around an issue. We then examine a case study of a funder, Development Innovations Ventures at USAID. Some of the innovations that it has funded were driven by social entrepreneurs without researcher involvement and some were tested using RCTs and/or had close involvement of development economics researchers. A review of this portfolio suggests that several programs involving development economics researchers and RCTs had substantial real-world influence.

c. Are RCTs that are more "academic" less useful for policy?

Many studies seek not to test just a particular program, but also to contribute to a body of literature that seeks to test different theories of human behavior. If citizens vote for candidates based on their ethnicity or caste is that because of very strong preferences, clientelistic networks, or a combinations of weak preferences and no alternative information on candidate quality? Do people only value what they pay for? How important are liquidity constraints, as opposed to lack of information or low human capital, in explaining poor child health and low business profitability in low-income families?

The studies that seek to answer these questions do not always test standard development programs, although some time they may turn into development ideas. De Mel, McKenzie, and Woodruff (2012) gave cash to businesses in Sri Lanka without conditions, repayment requirements, or mentoring, something unheard of in finance programs at the time (of course, eventually, the idea of unconditional cash transfers caught on as a realistic policy option, as indicated by the success of GiveDirectly). As we have discussed above, a series of studies that focused on pricing of health goods first asked households if they were willing to purchase a good at one price and then gave them the good at a lower price or for free, not something a regular program would do. Researchers pushed to test unconditional cash transfers (Baird et al., 2011; Blattman et al., 2014; Haushofer and Shapiro, 2013; Benhassine et al., 2014), even though at the time the political consensus was on conditional transfers.

The reason why this is potentially important for policy, and not just for academic curiosity, is that even where certain program specifics do not generalize, underlying patterns in human behavior may. The finding that small incentives are effective in encouraging people to take actions that have short-run costs but long-run benefits is more likely to generalize than the finding that lentils are a successful incentive for vaccination in Rajasthan (Banerjee and Duflo, 2010). Kremer and Glennerster (2011) review over 70 health economics RCTs and find strong similarities in consumer behavior across countries and products, including sharp reductions in take-up of non-acute care health products with small increases in price, big increases in take-up of non-acute products with small incentives (negative prices), and no evidence that paying for something makes people more likely to use it (Kremer and Miguel, 2007; Cohen and Dupas, 2010; Ashraf, Berry, and Shapiro, 2010a; Dupas, 2013).

This body of work on prices was taken up by advocates of free distribution of Insecticide Treated Bednets (ITNs). For many years there had been a fierce debate on the merits of free distribution, with free distribution advocates arguing that even small prices deter the poor, while others argued that small copayments were important to ensure ITNs were utilized. Armed with the evidence from RCTs, advocates of mass free distribution have successfully pushed this approach resulting in a dramatic rise in ITN coverage across Africa from roughly 2009 to 2015. The WHO reports that 43 of 47 countries in sub-Saharan Africa with ITN distribution programs provide them for free (World Malaria Report 2015). A recent article in Nature (Bhatt et al., 2015) examines the sharp decline in malaria infections in sub-Saharan Africa and estimates that between 2000 and 2015 malaria interventions prevented 663 million malaria cases, most of which is attributable to the sharp rise in ITN coverage: 450 million cases of malaria prevented by ITNs, and roughly 4 million deaths from 2000 to 2015.

Beyond the specific example of malaria, the policy community is coming to a more general realization that higher prices for preventive health products can sharply decrease take up and that price elasticity of demand can be very high (Kremer and Holla, 2009; Kremer and Glennerster, 2011; Dupas, 2016). This is changing the entire approach to pricing of these products.

Another area where a body of evidence from RCTs has produced both specific policy changes and given rise to more general lessons that have profoundly changed the policy debate is on attitudes toward cash transfer programs. Arguably the biggest innovation in anti-poverty and social protection policies in developing countries over the past twenty years is the growth of Conditional Cash Transfer programs (CCTs). Beginning in Mexico, these have now spread to more than thirty countries, and they have arguably played an important role in the decline in poverty in Latin America (Alzua, Cruces, and Ripani, 2013; Attanasio et al., 2005; Barrera-Osorio et al., 2011; Galiani and McEwan, 2013). While many factors were at play in the spread of Conditional Cash Transfers, we, along with many others, think the PROGRESA experiment (Schultz, 2004; Gertler, 2004) and the many following experiments in other contexts[4] played a significant role, through influencing Mexico's decision to continue and expand CCTs

---

[4] See Galiani and McEwan(2013), Glewwe and Olinto (2004), Benhassine et al (2014), Chaudhury, Friedman and Onishi (2013), Malucio and Flores (2005) among others, as well as a review in Fiszbein and Schady (2009)

after the inauguration of a new administration, the active promotion of CCTs by the Inter-American Development Bank and the World Bank, and the adoption of CCT by many countries.

More recently, there has been additional examination of how conditional cash transfers work that is further changing the policy debate. Conditional cash transfers have been shown by RCTs to not only increase the behavior on which the cash is conditional but to also improve outcomes such as height, weight, and cognitive development (Barham et al., 2013) and reduce HIV infection (Baird et al., 2011). There is also no evidence that poor households spend increased cash on alcohol or other temptation goods (Haushofer and Shapiro, 2013; Masterson and Lehmann, 2014; Evans and Popova, 2014). Indeed, the evidence suggests that the income elasticity of demand for food out of cash transfers is surprisingly high (see a review in Banerjee (2015)), and food transfers do not improve nutrition more than cash transfers (Cunha, 2014).

This is leading to a movement from a situation in which policymakers would almost never consider cash transfers to one in which cash transfers, conditional or not, are becoming an accepted tool in development policy. For example, as the world struggles to cope with refugees from war, groups such as the International Rescue Committee have drawn on RCTs of cash distributions in stable environments and with refugees (Masterson and Lehmann, 2014) to strongly push for cash rather than in-kind support for refugees. David Miliband, IRC president and CEO, said:

> The spate of man-made and natural disasters enveloping innocent civilians raises profound questions not just for international politics, but for NGOs and the humanitarian sector, as well. If we keep doing 'business as usual,' the gap between need and provision will continue to grow. Cash distribution – alongside clear humanitarian 'floor' targets in the revised Millennium Development Goals, more sustainable local partnerships and better use of evidence overall – could be part of a vital renewal of the humanitarian sector.

Early in the introduction of RCTs, Lant Pritchett (Pritchett, 2002) argued that RCTs would never become particularly popular with policymakers because they have reason to prefer ignorance over rigorous knowledge, in order to continued favoring their preferred program: "It pays to be ignorant." While in some cases policymakers may have incentives to preserve ignorance, in other cases policymakers are aware of the holes in their knowledge and would like to learn more. They may have a strong attachment to a favorite program, either due to inertia or a political imperative. But the experience of running the program often persuades them that they could do it better, and they are surprisingly open to ideas about how to improve their programs. The Raskin and MGNREGS programs mentioned above, where several teams of researchers have worked with the government, are good examples—while it was clear that the programs would continue, finding ways to make them work better was clearly of interest.

b. How to assess the policy success (or not) of the RCT agenda

It is a little bit difficult to assess the causal effect of RCTs on policy adoption. Interventions subject to RCTs are not themselves randomized, and many factors influence whether and when a particular intervention is taken up. When a program is taken up after an RCT showed it has worked, it is not always because of the RCT, and it is never *just* because of the RCT. Nevertheless, some have argued that the influence of RCTs on policy is actually quite low, compared to the volume of RCTs. For example, Shah, Wang, Nadel and Fraker (2015) point out that despite the 489 completed J-PAL evaluations, there are only 9 scale-up or policy influence stories on J-PAL's web site. But this number *per se* is not particularly informative: for one, it is not a census of the studies that have some impact. Not all J-PAL RCTs are systematically followed up. These stories are chosen precisely because of the size of their impact and because they can be documented clearly. The absolute number of lives reached by them is quite significant—the JPAL website tells us that over 200 million people were reached by these programs. But the main concerns with any statistic like this are conceptual:

1. The J-PAL web-site does not carry statistics on non-J-PAL studies for the very good reason that, based on our experience collecting information from DIV and J-PAL, it is far from straightforward to collect information on the extent to which RCTs have influenced policy. This means, for example, that the number does not include the hundreds of millions of people who have been reached by CCT.

2. Many RCTs are fairly recent. Taking these to the policy level requires a lot of care, especially given the external validity issues (Would it work in government? Would it work in a different place?) The process is therefore often slow, again for very good reasons. Therefore, we should not expect a lot of these to be scaled as yet.

3. Many of the most valuable RCTs are those that test popular and highly touted policies that already exist in the world on a large-scale and show that they are in fact much less effective than previously claimed or believed. Microfinance and improved cook-stoves are two obvious examples. In such cases, success would be to slow down the spread of such policies. Obviously, in such cases, one would not expect something to appear on the JPAL scale-up page, but these are two cases where the work has probably been quite influential.

4. In some cases, the primary purpose of an RCT is not to directly affect policy, but to instead investigate an underlying theoretical mechanism, which may, in turn, indirectly influence policy. However, such cases would not appear on a list of scale-ups, despite the fact that the knowledge they have provided has impacted, albeit indirectly, a large number of people. For example, the orthodoxy in development economics had long been that the poor are "poor but rational." The accumulating evidence from RCTs has undoubtedly hastened the diffusion of the idea into development economics and into development policy that poor people are not always rational: this is reflected for example, in both the content, and the number of RCTs in the World Development Report (2016) on psychology and poverty. In turn, publications like the WDR, and the associated discussions, influence the design of policies.

5. It is not clear what the right benchmark for success should be. We suspect that if one looked at other areas of economics, one would find that research projects influenced policy at a much lower rate than RCTs have in development policy in recent years. Moreover, one would not want to say that rapid policy influence is the sole or even the major metric by which the worth of economic research should be assessed—think of the idea of congestion pricing for road use (Vickrey, 1969), which is only beginning to find real world applications.

6. Perhaps most importantly, it is worth realizing that the payoff to RCTs is likely to be the average of a highly skewed distribution. Looking at the fraction of RCTs which scale, rather than the average payoff, is therefore as misleading as looking at the fraction of any research and development effort which succeeds in terms of say generating a successful marketed product, since the payoff to research and development in general is typically very highly skewed. As is well-known, citations across scientific disciplines appear to follow a power law distribution, with a small fraction of papers accounting for the majority of citations. This peak is followed by a steep decay, as a large portion of research papers are never cited (Radicchi, Fortunato, and Castellano, 2008).[5] As we mentioned, the 9 policy innovations listed by J-PAL together have reached over 200 million people, and this does not include the more than 100 million people who have been reached through India's most recent round of deworming, the millions of people who have received free bed nets (since J-PAL lists it as policy influence but does not provide a count), and the 60 million people whose water and air is less polluted because of the statewide adoption of better regulation of industrial pollution in Gujarat (again, not counted).

7. For this reason, pointing out that many research and development efforts yield low payoffs does not suggest that these are bad investments ex ante. The correct analytical question to ask is whether the expected average or marginal payoff to R&D effort in RCTs is positive or greater than that in other areas of research if one takes overall research budgets as fixed. Of course, measuring the payoff to research is inherently a difficult exercise for all sorts of conceptual reasons. There is also the added statistical difficulty that a large amount of data is needed to accurately measure the mean of a fat-tailed distribution.

---

[5] For instance, in the social sciences in general, papers receive on average 0.5 citations in the first two years of publication, including self-citations (Klamer and Dalen, 2002), whereas in mathematics, medicine, and education the number is estimated to be less than 1 (Mansilla et al. 2007). Given the skewed distribution, this implies that the median paper is never cited. Similarly, the majority of new patents have extremely low value with a small fraction of patents accounting for much of the overall value of patents.

c. What have we learnt from the DIV experience?

Keeping all of this in mind, we now turn to one particular example, the experience of the investments made by USAID's Development Innovation Ventures (DIV) between 2010 and 2012.

DIV holds a year-round grant competition for innovative solutions to a range of development challenges, pilots and tests them using analytical methods, and scales solutions that demonstrate widespread impact and cost-effectiveness. DIV supports novel business or organizational models; operational, behavioral or production processes; and products or services that can help address development challenges. DIV's tiered-funding model, provides small grants to pilot innovations in development; medium-size grants to rigorously test for impact and cost-effectiveness (often using RCTs) or ability to pass a market test; and larger-scale grants to help transition innovations to scale that have passed a market test or that have rigorous evidence of impact and cost-effectiveness.

When DIV was established, two targets were set for the program: 1) a 15% social rate of return on investment, and 2) a reach of at least 75 million people worldwide, through direct investment and through broader influence on the rest of USAID. Preliminary work by DIV staff suggests that the 2010-2 portfolio easily met the first goal, even under the conservative assumptions that all innovations supported by DIV yielded no further benefits, and even looking only at a subset of innovations that yielded financial benefits or health benefits that could be valued in terms of DALYs. While social return is a more conceptually comprehensive measure for evaluating DIV, it is difficult to measure, and this piece seeks not to evaluate DIV, but rather to look at the narrower question of whether RCTs can have real world influence. We therefore focus on examining the number of people reached by innovations supported by DIV (as well as by later adapted versions of these innovations). (Note that substantial reach is a necessary but not sufficient condition for high social return, since the total social benefit of an innovation equals the net benefit per person reached times the number of people reached.) This exercise is inherently limited, so readers will have to make their own judgements about the likely impact per person reached, the likely future reach of these innovations (sustainability), and the extent to which DIV funding played an important role in the reach achieved by innovations in the DIV portfolio. What we are doing here is rather the descriptive exercise of systematically tracking a portfolio. Nevertheless, following the entire 2010-2012 DIV portfolio is interesting for a paper that explores the influence of RCTs, because the premise of DIV is specifically to fund innovations in development that have the potential to cost-effectively reach a large number of people through either the public or the private sector.

In particular, whereas many other programs have a top-down approach in which program staff identify problems in advance, choose sectors on which to focus, or set strategy within sectors, DIV follows a bottom-up approach that is deliberately open across sectors: supporting innovations that will scale commercially, innovations designed to scale through the public sector, and startups and organizations proposing to change behavior within existing large organizations. Although the bulk of DIV's outreach effort has been oriented towards traditional social entrepreneurs, DIV has also made an effort to be open to proposals from development economics researchers. To balance this

openness, DIV employs a staged finance approach in which innovations only receive larger-scale support after they have passed rigorous tests. DIV provides large-scale support (Stage 3) only for innovations which have rigorous evidence of impact and cost-effectiveness or which have demonstrated market viability. At the piloting (Stage 1) and testing stages (Stage 2), however, DIV has historically been open to proposals that have the *potential* to scale based on their cost-effectiveness, for example, even if they do not necessarily already have a management team in place capable of scaling internally or written commitments from scaling partners.[6]

This combination of approaches thus helps us ask whether the engagement with the development economics research community, and the willingness to consider early-stage investments even without a fully proven capacity to scale, came at the cost of scaling success. We can shed light on these questions by comparing the scaling record across types of projects, stages of funding, and of course by looking at the scaling record of DIV.

In the Appendices, we provide a list of all the DIV awards from this time period, and a narrative description of the innovations that have, subsequent to DIV's funding, reached more than 100,000 people.

Table 3 shows the results of this exercise. Here are some key insights:

1) **DIV has been relatively successful in supporting innovations that scale. A relatively high fraction of DIV awards, and an even higher fraction of DIV total investment, went to projects which, to date, have already reached more than 100,000 people (and a smaller but still high fraction of the awards went to projects that reached more than a million people).**

   30% of DIV awards (13/43) have so far reached more than 100,000 people within 3-5 years.[7] These awards account for 57% of the total value of DIV awards in this time period, or $10.98 million in total funding. 14% of DIV awards (6/43) have so far reached more than one million people. These awards account for 33% of the total value of DIV awards in this time period, or $6.38 million in total funding.

---

[6] Although DIV does not require a proven pathway to scale at Stages 1 or 2, a promising pathway to scale through the public or private sector (or a hybrid of the two) and strong potential demand is one of its main selection criteria, particularly at Stage 2.

[7] Two innovations (that reached over 100,000 people) received both a Stage 1 and a Stage 2 award. Thus, these 12 awards support 10 separate innovations.

Why do we say that 30% is "relatively successful"? A rule of thumb in the venture capital world is that 10% of investments yield modest success and 1% yield large successes. While we have not yet identified other funders that publish data that would allow for computation of comparable statistics, our reading of the literature and our examination of websites of some other organizations suggests that these rates compare well with those achieved over a much longer time frame by other impact-investing organizations. These results are all the more striking because, while some organizations provide funding only after a certain level of scale is reached (e.g., Acumen, Skoll Foundation), DIV often supported innovations at an early stage (as well as tests to know whether they were worth scaling up), rather than waiting until innovations had already reached a certain scale and had attracted earlier support before investing.

2) **Stage One and Stage Two awards have a particularly low DIV expenditure per person reached and account for more than 90% of people reached by innovations supported by DIV during this time period.**

One of these early stage innovations (Consumer Action and Matatu Safety) recently received a Stage 3 DIV award, but in general, Stage 1 and 2 innovations attained high levels of reach because other funders/entities provided support based in part on the information generated from the DIV-funded project.

3) **While the estimated DIV expenditure per person is lower for earlier stage grants, it is fairly low across the board. This is because the great majority of the reach of DIV-supported innovations was attained without the applicants returning to DIV for additional financial support.**

Though many past-awardees apply for additional funding, only 7% of DIV's 2010-2012 portfolio of grantees received follow-on funding after the initial period of performance. Over 40% of DIV's 2010-2012 grantees received follow-on funding from either the public or private sector after DIV's investment. DIV's capacity to be catalytic of course partly derives from the rich funding ecosystem in which it operates, where other entities (governments, NGO, private sector firms) can adopt innovations.

4) **Cost was a key determinant of which innovations scaled. The largest scale was achieved by innovations with very low costs per person.**

   In some cases, the innovations involved the provision of information by media or phone (e.g., voter report cards, election monitoring), or provided behavioral "nudges" in large, existing systems (e.g., Zambian community health workers). Of course it's important to recognize that total impact depends on the benefit per person reached times the number of people reached, and some innovations with moderate cost per person (e.g. VisionSpring) and moderate reach may generate high total social benefit because the benefit per person is very high.

5) **While some innovations reached more than 100,000, or in one case, more than 1,000,000 people through the creation and growth of a new organization designed to scale the innovation, the vast majority of reach was delivered through adoption by existing large organizations, including large firms, NGOs, and governments.**

   Four of the DIV-supported innovations which reached 100,000 or more consumers involved the creation of new organizations which scaled from scratch. Seven involved adoption of the innovation by existing entities that already had high levels of reach.

   Of the six innovations which reached more than one million people, one was scaled by an NGO which constructed and built operations around the innovation (Evidence Action in the case of chlorine dispensers), and four did so by adoption by existing organizations (an insurance company and the Kenyan National Transport and Safety Authority in the case of stickers in matatus, the Government of India in the case of biometric monitoring, political campaigns in the case of real-time efforts to send polling station outcomes to central locations by mobile phones, and newspapers in the case of voter report cards). Existing organizations with large reach that adopted DIV-supported innovations or modified versions of these innovations included private sector firms, NGOs, and governments.

6) **Innovations tested with RCTs scale not only through adoption by governments, but also through adoption by private sector firms and NGOs.**

Of the ten DIV awards for innovations with RCTs that have reached more than 100,000 people, there were two clear cases in which developing country governments played the lead role (i.e. scaling of an improved approach to community health worker recruitment by the government of Zambia and biometric monitoring in India). The Kenyan government seems likely to play an important role alongside the insurance industry in scaling the Kenyan matatu safety program. Donors played a key role in provision of Potential Energy's improved cookstoves in Darfur. NGO partners played a role in a number of projects. A major lesson of this analysis is that large private firms played a major role as well (e.g., an insurance company played a key role in the matatu stickers project and newspapers published the free content when an NGO provided them with voter reports cards).

7) **Innovations involving randomized controlled trials, and/or developed in part by researchers (often working in close conjunction with implementers), reach 100,000 or 1,000,000 users at a particularly high rate.**

43% (10/23)[8] of awards for which an RCT was used for evaluation or development economics researchers were involved in design of the innovation reached more than 100,000 people.[9] 26% (6/23) of these awards supported innovations that had reached more than one million people in the original or adapted form (e.g. voter report cards, election monitoring, stickers in matatus, chlorine dispensers, and biometric attendance verification). In contrast, among the innovations not including an RCT component or a strong role for development economics researchers, only 16% (3/19) reached 100,000 people (Vision Spring, Mera Gao, d.light), and none reached more than one million people.[10]

---

[8] Projects were coded as having development economics researchers involved if the initial proposal that was funded by DIV explicitly included the efforts of researchers. Although d.light's initial proposal included an RCT on the impacts of their products, this RCT did not take place and funding strictly supported the development of a new solar home system as well as an *ex post* impact evaluation of these systems. Due to these circumstances, we have not included d.light in our calculation of projects developed in part by researchers in this point. If we were to include d.light, this figure would be 11/24, or 46%.

[9] Voter information report cards (2 awards), election monitoring technology, digital attendance and medical information systems in primary health care centers, mobile tools for community health care workers (2 awards), consumer action on Matatu safety, bringing safe water to scale, improved cookstoves, and recruiting community health workers.

[10] 24 awards incorporated an RCT component or were based on an RCT. This excludes two cases in which the initial proposal included an RCT but the ultimate actual project funded by DIV did not include an RCT: Psychometric Analysis for Entrepreneurs (AID-OAA-F-13-00028) and Affordable Access to Energy for All: Innovative Financing for Solar Systems (AID-OAA-F-13-00007). Note that since there is a lot of overlap between researcher-led projects and projects with an RCT, we cannot easily separate their impact.

One could imagine multiple hypotheses for this difference in the rates of success. First, it might be easier to reach many people by persuading large organizations and governments to adopt the innovation and in this process the evidence from the RCTs might have played an important role. By contrast those innovations that did not come from the academic RCT side tried to scale by directly implementing or selling their product, which may be harder, as these innovations do not have large pre-existing policies, programs or institutions as initial partners. Second, it is often argued that academic researchers mainly want to publish, and this conflicts with their incentives to get involved in projects that are socially useful but not as creative (replication, tinkering with design, etc.). But on the other hand, it is also argued that journals have a strong publication bias, and it is easier to publish things that have worked. Ergo, development economists should have strong incentives to develop and test innovations that have a reasonable chance of success. Moreover, perhaps (just perhaps?) economics actually gives them some useful insights into the design of projects. Third, it may also be that the recent focus on information and behavioral economics makes them particularly interested in innovations with a low cost per user ("nudges"), which seems to be a strong predictor of success. Fourth, when researchers were involved, they were typically not just evaluators: they were fully involved in the development of the innovation (e.g. voter report cards, chlorine dispensers, monitoring project in Afghanistan), worked closely with implementing organizations, and remained closely involved in the details of the implementation. They were in fact "researcher-entrepreneurs". Many of the ideas developed by researchers drew on the latest ideas in the field, and the data suggest that the researchers who developed these ideas were then relatively successful in working with others to scale these innovations.

8) **Innovations that had already been tested through RCTs and found to have impact and potential for cost-effectiveness prior to applying for DIV support accounted for three of the five innovations that reached over one million people.**

Three of the five innovations that reached over one million people (voter report cards, Consumer Action and Matatu Safety, and Chlorine Dispensers for Safe Water) had already been subject to RCTs before applications were submitted to DIV. While we have not yet coded the data, we believe that there were very few applications in this category, so the rate at which proposals in this category reached over one million people was very high (possibly 100%).

9) **While some DIV-supported innovations have been applied in multiple countries, most have not.**

So far, DIV-supported innovations have typically not been applied much beyond the country where they have been tested. This may be an area where future work is needed.

Conclusion

The discussion in this policy section suggests that RCTs have influenced policy both through providing evidence on individual projects and programs, and by changing thinking in development more broadly.

The biotech and IT industries routinely build on innovations developed by researchers using frontier techniques in those fields. The evidence from DIV awards is consistent with the idea that a similar approach may be effective in development, with innovations developed in part by researchers and/or involving randomized controlled trials reaching 100,000 or 1,000,000 users at a particularly high rate. This is absolutely not to say that work is not needed to fine tune interventions for different contexts, or that it is not important to evaluate real world programs that have not yet been evaluated using an RCT; but the development of new ideas that are grounded in basic science actually can lead to real-life change.

One striking lesson of this analysis is that the projects that are scaled up tend to be low-cost, well-defined, and simple. Other examples, not in this list, also fit this bill (e.g., deworming, the Raskin card). There are notable counterexamples of programs that are neither particularly cheap nor simple and have scaled up: Conditional Cash Transfers and the BRAC ultra poor programs are two examples. Furthermore, those two programs were not only scaled up where they had been tested, but were implemented in many other countries as well. Interestingly, they were initially replicated as RCTs.

Well-defined interventions are also the ones that are more likely to lead to successful research projects, since they can more easily pin down a specific mechanism, and be construed as a test for a theory. So the reasons why RCTs have been so successful as a research tool may also be what makes them successful at leading to real world changes.

Looking forward, we don't know what the most important pathways of influence for RCTs might turn out to be. One route is that simple, clear insights, low-cost interventions, or low-cost modification to promising existing programs get adopted, as the DIV case study suggests. The fact that these innovations are low-cost of course does not mean that they have low impact. One lesson from decades of well-identified development research is that details are incredibly important, and that the distinction between "big" and "small" questions can be very misleading (see Banerjee and Duflo (2012), Chapter 10 for a longer discussion).

An alternative pathway is one in which more complex interventions are replicated in many contexts and then widely adopted, following the PROGRESA or the BRAC model. The third one is that rather than just focusing only on the results, policymakers and other actors adopt the experimental attitude, i.e. leave some

space within their operations for innovations and learning perhaps housed inside a specialized unit (like the White House "nudge" unit) or a cross-department fund (like the Tamil Nadu innovation fund).

But to really get the full benefits of the RCT revolution, it is not enough to do more RCTs and get some of them scaled up. A range of complementary institutions are also necessary to more effectively translate research into policy. For example, we need better systems for the production of meta-analyses and review articles and for the creation of expert panels to review the evidence. Medicine has a quite involved system for this, but even setting aside the question of how well that system works in medicine (Sim, 2001; Kawamoto et al., 2005), the institutions that are appropriate for medicine are not necessarily those that are appropriate for social science, and development economics in particular. These institutions are just starting to get built: the AEA registry of RCTs is an example of a successful effort to build a registration platform. Its popularity suggests that the development community is receptive to these efforts.

In addition to the purely scientific infrastructure for learning, the process of going from an idea to a program at scale requires appropriate institutional support. First, funders are needed to finance iterative piloting before an RCT to work out the implementation details.[11] Once an RCT has been conducted, institutional support is also needed for iterating on the intervention to "ready" it to transition to scale. This includes testing ways to: bring unit costs down (since the first RCT often evaluates a small pilot with high unit costs); collaborate with potential implementing partners; and mitigate potential cost increases and/or reduced benefits that may results from institutional and personnel differences between the pilot and scaled-up versions of an innovation (due to, for example, government procurement systems with higher transaction costs or limited government capacity to implement the intervention effectively). To get to the right scaled-up version therefore involves trying them out to scale and measuring the impact at scale. Indeed, multiple iterations may be needed until something that is appropriate for policy can work. Figuring out how best to do the scaling in each case or how to do so in additional countries takes time, specialized human capital and additional funding.

---

[11] Development Innovation Ventures and the Global Innovation Fund – a private fund modeled after DIV and to which DIV and other bilateral donors and impact investors contribute – explicitly encompass such a piloting phase.

References:

Abdulkadiroglu, Atila, Joshua Angrist, and Parag Pathak. "The elite illusion: Achievement effects at boston and new york exam schools." *Econometrica* 82, no. 1 (2014): 137-196.

Allcott, Hunt. "Site Selection Bias in Program Evaluation." *The Quarterly Journal of Economics* (2015): qjv015.

Alesina, Alberto, Paola Giuliano, and Nathan Nunn. "On the Origins of Gender Roles: Women and the Plough." *The Quarterly Journal of Economics* 128.2 (2013): 469-530.

Ashraf N., Berry, J., & Shapiro, J. (2010a) "Can higher prices stimulate product use? Evidence from a field experiment in Zambia." *American Economic Review*. (in press).

Ashraf, Nava, Oriana Bandiera, and Scott S. Lee. "Do-gooders and Go-getters: Career Incentives, Selection, and Performance in Public Service Delivery." *Working Paper*, March 2015.

Ashraf, Nava, James Berry and Jesse M. Shapiro. 2010. "Can Higher Prices Stimulate Product Use? Evidence from a Field Experiment in Zambia." *American Economic Review*, 100(5): 2383-2413.

Ashraf, Nava, Dean Karlan, and Wesley Yin. "Tying Odysseus to the mast: Evidence from a commitment savings product in the Philippines." *The Quarterly Journal of Economics* (2006): 635-672.

Athey, Susan, and Guido Imbens (2016). "The Econometrics of Randomized Experiments," Forthcoming, *Handbook of Field Experiments*, Eds. Abhijit Banerjee and Esther Duflo.

Attanasio, Orazio, Erich Battistin, Emla Fitzsimons, and Marcos Vera-Hernandez. 2005. "How effective are conditional cash transfers? Evidence from Colombia." *Institute for Fiscal Studies Briefing* Note.

Attanasio, Orazio, Camila Fernández Emla O A Fitzsimons, Sally M Grantham-McGregor, Costas Meghir, Marta Rubio-Codina, Using the infrastructure of a conditional cash transfer program to deliver a scalable integrated early child development program in Colombia: cluster randomized controlled trial BMJ 2014;349:g5785

Attanasio, Orazio, Arlen Medina and Costas Meghir 2016 "Long Term Impact of Vouchers for Vocational Training: Experimental Evidence for Colombia, Forthcoming, American Economic Journal, Applied Economics

Attanasio, Orazio, Costas Meghir and Ana Santiago (2011) "Education Choices in Mexico: Using a Structural Model and a Randomized Experiment to evaluate PROGRESA," Review of Economic Studies(2011)doi: 10.1093/restud/rdr015

Baird, Sarah, Craig McIntosh, and Berk Özler. "Cash or condition? Evidence from a cash transfer experiment." *The Quarterly Journal of Economics*(2011): qjr032.

Bandiera, O., R. Burgess, N. Das, S. Gulesci, I. Rasul, and M. Sulaiman. 2013. Can Basic Entrepreneurship Transform the Economic Lives of the Poor? IZA Discussion Papers 7386. Bonn: Institute for the Study of Labor (IZA).

Banerjee, Abhijit Vinayak, et al. "Making Aid Work." MIT Press Books 1 (2007).

Banerjee, Abhijit, Jishnu Das and Reshmaan Hussam. 2016. "Improving the Quality of Private Sector Health Care in West Bengal" Forthcoming.

Banerjee, Abhijit Vinayak. "Big answers for big questions: the presumption of growth policy." *Brookings Conference What Works in Development*. 2008.

Banerjee, Abhijit V. "Policies for a Better-fed World." Review of World Economics (2015): 1-15.

Banerjee, Abhijit V., and Esther Duflo. "The Experimental Approach to Development Economics." *Annu. Rev. Econ.* 1.1 (2009): 151-178.

Banerjee, Abhijit, et al. "Improving police performance in Rajasthan, India: Experimental evidence on incentives, managerial autonomy and training." No. w17912. *National Bureau of Economic Research*, 2012.

Banerjee, A., Duflo, E., Goldberg, N., Karlan, D., Osei, R., Parienté, W., ... & Udry, C. (2015). "A multifaceted program causes lasting progress for the very poor: Evidence from six countries." *Science*, *348*(6236), 1260799.

Banerjee, Abhijit V., and Esther Duflo. "Do firms want to borrow more? Testing credit constraints using a directed lending program." The Review of Economic Studies 81.2 (2014): 572-607.

Banerjee, Abhijit, Esther Duflo, Clément Imbert, Santhosh Mathew, and Rohini Pande. (2016) "Can e-governance reduce capture of public programs? Experimental evidence from India's Employment Guarantee." *Mimeo.*

Banerjee, Abhijit, Sylvain Chassang, and Erik Snowberg. "Decision Theoretic Approaches to Experiment Design and External Validity." No. w22167. *National Bureau of Economic Research*, 2016.

Banerjee, Abhijit, Rema Hanna, Jordan C. Kyle, Benjamin A. Olken, and Sudarno Sumarto. "The Power of Transparency: Information, Identification Cards and Food Subsidy Programs in Indonesia." No. w20923. *National Bureau of Economic Research*, 2015.

Barham, Tania, Karen Macours, and John A. Maluccio. *More schooling and more learning? Effects of a three-year conditional cash transfer program in Nicaragua after 10 years*. No. IDB-WP-432. IDB Working Paper Series, 2013.

Barrera-Osorio, Felipe, Marianne Bertrand, Leigh L. Linden, and Francisco Perez-Calle. 2011. "Improving the Design of Conditional Transfer Programs: Evidence from a Randomized Education Experiment in Colombia." *American Economic Journal: Applied Economics*, 3(2): 167-95.

Basu, Kaushik. "Randomisation, Causality and the Role of Reasoned Intuition." *Oxford Development Studies* 42, no. 4 (2014): 455-472.

Benhassine, Najy, Florencia Devoto, Esther Duflo, Pascaline Dupas, and Victor Pouliquen. 2015. "Turning a Shove into a Nudge? A 'Labeled Cash Transfer' for Education." *American Economic Journal: Economic Policy, Forthcoming*.

Bertrand, Marianne, Simeon Djankov, Rema Hanna, and Sendhil Mullainathan. "Obtaining a driver's license in India: an experimental approach to studying corruption." *The Quarterly Journal of Economics* (2007): 1639-1676.

Bertrand, Marianne, and Esther Duflo. "Field experiments on discrimination." No. w22014. *National Bureau of Economic Research*, 2016.

Bertrand, Marianne, and Sendhil Mullainathan. "Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination." *American Economic Review* 94, no. 991 (2004): 1013.

Bhatt, S., et al. "The effect of malaria control on Plasmodium falciparum in Africa between 2000 and 2015." *Nature* 526.7572 (2015): 207-211.

Blattman, Christopher, Julian C. Jamison, and Margaret Sheridan. "Reducing crime and violence: Experimental evidence on adult noncognitive investments in Liberia." No. w21204. *National Bureau of Economic Research*, 2015.

Blattman, Christopher, Nathan Fiala, and Sebastian Martinez. 2014. "Generating skilled employment in developing countries: Experimental evidence from Uganda." *Quarterly Journal of Economics,* 129(2): 697–752.

Bolton, Paul, Judith Bass, Theresa Betancourt, Liesbeth Speelman, Grace Onyango, Kathleen F. Clougherty, Richard Neugebauer, Laura Murray, and Helen Verdeli. 2007. "Interventions for Depression Symptoms Among Adolescent Survivors of War and Displacement in Northern Uganda: A Randomized Controlled Trial." *JAMA* 298 (5): 519. doi:10.1001/jama.298.5.519.

Bolton, Paul, Judith Bass, Richard Neugebauer, Helen Verdeli, Kathleen F Clougherty, Priya Wickramaratne, Liesbeth Speelman, Lincoln Ndogoni, and Myrna Weissman. 2003. "Group Interpersonal Psychotherapy for Depression in Rural Uganda: A Randomized Controlled Trial." *Jama* 289 (23): 3117–24.

Bursztyn, Leonardo, and Lucas C. Coffman. 2012. "The Schooling Decision: Family Preferences, Intergenerational Conflict, and Moral Hazard in the Brazilian Favelas." *Journal of Political Economy* 120 (3): 359–97. doi:10.1086/666746.

Callen, Michael, and James D. Long. "Institutional corruption and election fraud: Evidence from a field experiment in Afghanistan." *The American Economic Review* 105, no. 1 (2014): 354-381.

Cameron, D. B., Mishra, A., & Brown, A. N. (2015). The growth of impact evaluation for international development: how much have we learned?.*Journal of Development Effectiveness*, 1-21.

Chaudhury, Nazmul, Jed Friedman, and Junko Onishi. 2013. "Philippines conditional cash transfer program impact evaluation 2012." Manila: World Bank Report Number 75533-PH.

Chassang, Sylvain, Gerard Padró i Miquel, Erik Snowberg "Selective trials: A principal-agent approach to randomized controlled experiments." *American Economic Review* 102 (2012): 1279-1309.

Clark, Damon. "The performance and competitive effects of school autonomy." Journal of political Economy 117.4 (2009): 745-783.

Cohen, Jessica, and Pascaline Dupas (2010). "Free Distribution or Cost-Sharing? Evidence from a Randomized Malaria Prevention Experiment" *Quarterly Journal of Economics*, 125(1), pp. 1-45.

Congdon, William J., Jeffrey R. Kling, Jens Ludwig, and Sendhil Mullainathan (2016). "Social Policy: Mechanism Experiments & Policy Evaluations," Forthcoming, *Handbook of Field Experiments*, Eds. Abhijit Banerjee and Esther Duflo

Cunha, Jesse M. "Testing Paternalism: Cash versus In-Kind Transfers: Dataset." *American Economic Journal: Applied Economics* (2014).

Dal Bó, Ernesto, Frederico Finan, and Martín A. Rossi. "Strengthening state capabilities: The role of financial incentives in the call to public service." *The Quarterly Journal of Economics* 128, no. 3 (2013): 1169-1218.

Deaton, Angus. "Instruments, randomization, and learning about development." *Journal of economic literature* (2010): 424-455.

Dell, Melissa. "The persistent effects of Peru's mining mita." *Econometrica* 78, no. 6 (2010): 1863-1903.

De Mel, Suresh, David McKenzie, and Christopher Woodruff. "Returns to capital in microenterprises: evidence from a field experiment." *The Quarterly Journal of Economics* (2008): 1329-1372.

De Mel, Suresh, David McKenzie, and Christopher Woodruff. "One-time transfers of cash or capital have long-lasting effects on microenterprises in Sri Lanka." Science 335.6071 (2012): 962-966.

Digital Dividends. Washington, DC: International Bank for Reconstruction and Development The World Bank, 2016. *World Bank*. Web. 31 May 2016. <http://www.worldbank.org/en/publication/wdr2016>.

Dobbie, Will, and Roland G. Fryer Jr. "The impact of attending a school with high-achieving peers: evidence from the New York City exam schools." *American Economic Journal: Applied Economics* 6.3 (2014): 58-75.

Duflo, Esther "Scaling up and Evaluation" (2004) ABCDE Annual World Bank Conference on development economics ,

Duflo, Esther & Michael Kremer. 2005. "Use of Randomization in the Evaluation of Development Effectiveness." in George Pitman, Osvaldo Feinstein and Gregory Ingram (eds.) Evaluating Development Effectiveness. New Brunswick, NJ: Transaction Publishers.

Duflo, Esther, Rachel Glennerster, and Michael Kremer. "Using randomization in development economics research: A toolkit." *Handbook of development economics* 4 (2007): 3895-3962.

Duflo, Esther, Michael Greenstone, Rohini Pande, and Nicholas Ryan. "What does reputation buy? differentiation in a market for third-party auditors." *The American Economic Review* 103, no. 3 (2013a): 314-319.

Duflo, Esther, Michael Greenstone, Rohini Pande, and Nicholas Ryan. "Truth-telling by third-party auditors and the response of polluting firms: experimental evidence from India." *Quarterly Journal of Economics* 128, no. 4 (2013b): 1499-1545.

Duflo, Esther, Michael Kremer and Jonathan Robinson (2008a). "How High are Rates of Return to Fertilizer? Evidence from Field Experiments in Kenya." *American Economic Review Papers (Papers and Proceedings Issue)* 98 (2): 482–488

Duflo, Esther, Michael Kremer, and Jonathan Robinson. "Nudging Farmers to Use Fertilizer: Evidence from Kenya." *American Economic Review 101(6),* 2008b.

Duflo, Esther, Michael Kremer, Jonathan Robinson and Frank Schilbach. "Technology Diffusion and Appropriate Use: Evidence from Western Kenya". *Working paper, in progress.*

Dupas, Pascaline. "Short-run subsidies and long-run adoption of new health products: Evidence from a field experiment." *Econometrica* 82.1 (2014): 197-228.

Dupas, Pascaline. "Short-run subsidies and long-run adoption of new health products: Evidence from a field experiment." *Econometrica* 82, no. 1 (2014): 197-228.

Dupas, Pascaline, and Ted Miguel (2016). "Impacts and Determinants of Health Levels in Low-Income Countries," Forthcoming, *Handbook of Field Experiments*, Eds. Abhijit Banerjee and Esther Duflo

Dupas, Pascaline, and Jonathan Robinson. "The Daily Grind: Cash Needs, Labor Supply and Self-Control." *Unpublished paper* (2014).

Durden, Garey C., and Larry V. Ellis. "A method for identifying the most influential articles in an academic discipline." *Atlantic Economic Journal* 21.4 (1993): 1-10.

Dustan, Andrew, Alain de Janvry, and Elisabeth Sadoulet. "Flourish or Fail? The Risky Reward of Elite High School Admission in Mexico City." Working Paper. (2015).

Evans, David K., and Anna Popova. "Cash transfers and temptation goods: a review of global evidence." World Bank Policy Research Working Paper 6886 (2014).

Fisher, Ronald Aylmer. "Statistical methods for research workers." *Genesis Publishing Pvt Ltd*, 1925.

Finan, Frederico, Benjamin A. Olken, and Rohini Pande. "The Personnel Economics of the State." No. w21825. *National Bureau of Economic Research*, 2015.

Fiszbein, Ariel, and Norbert Rüdiger Schady. 2009. *Conditional Cash Transfers: Reducing Present and Future Poverty*. A World Bank Policy Research Report 47603. Washington D.C: World Bank.

Freedman, David A. "Statistical Models for Causation What Inferential Leverage Do They Provide?." *Evaluation review* 30, no. 6 (2006): 691-713.

Galiani, Sebastian, and Patrick J. McEwan. 2013. "The heterogeneous impact of conditional cash transfers." Journal of Public Economics 103: 85-96.

Gertler, Paul. "Do conditional cash transfers improve child health? Evidence from PROGRESA's control randomized experiment." *The American Economic Review* 94.2 (2004): 336-341.

Giné, Xavier, Dean Karlan, and Jonathan Zinman. "Put your money where your butt is: a commitment contract for smoking cessation." *American Economic Journal: Applied Economics* (2010): 213-235.

Glennerster, Rachel. (2016). "The Practicalities of Running Randomized Evaluations: Partnerships, Measurement, Ethics, and Transparency" Forthcoming, *Handbook of Field Experiments*, Eds. Abhijit Banerjee and Esther Duflo.

Glewwe, Paul, and Pedro Olinto. 2004. "Evaluating the Impact of Conditional Cash Transfers on Schooling: An Experimental Analysis of Honduras' PRAF Program." University of Minnesota Unpublished Manuscript.

Glewwe, Paul, Michael Kremer, and Sylvie Moulin. "Many children left behind? Textbooks and test scores in Kenya." *American Economic Journal: Applied Economics* 1.1 (2009): 112-135.

Gueron, Judy (2016). "The Politics and Practice of Social Experiments: Seeds of a Revolution" Forthcoming, *Handbook of Field Experiments*, Eds. Abhijit Banerjee and Esther Duflo.

Haushofer, Johannes, and Jeremy Shapiro. "Household response to income changes: Evidence from an unconditional cash transfer program in Kenya." *Massachusetts Institute of Technology* (2013).

Hanna, Rena and Dean Karlan (2016). "Designing Social Protection Programs: Using Theory and Experimentation to Understand how to Help Combat Poverty" Forthcoming, *Handbook of Field Experiments*, Eds. Abhijit Banerjee and Esther Duflo.

Heckman, James J. 1992. Randomization and Social Policy Evaluation. Harvard University Press. Pages 201—230.

Holla, Alaka and Michael Kremer. (2009). "Pricing and Access: Lessons from Randomized Evaluations in Education and Health." In *What Works in Development? Thinking Big and Small.* Eds. Jessica Cohen and William Easterly. Washington, DC: Brookings Institution Press, 2009.

de Janvry, Alain, Elisabeth Sadoulet and Tavneet Suri. (2016) "Agriculture in Developing Economies" Forthcoming, *Handbook of Field Experiments*, Eds. Abhijit Banerjee and Esther Duflo

Kawamoto, Kensaku, et al. "Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success." Bmj 330.7494 (2005): 765.

Klamer, Arjo & Hendrik P. van Dalen (2002) Attention and the art of scientific publishing, *Journal of Economic Methodology*, 9:3, 289-315, DOI: 10.1080/1350178022000015104

Karlan, Dean S., and Jonathan Zinman. "Credit elasticities in less-developed economies: Implications for microfinance." *The American Economic Review*(2008): 1040-1068.

Kreindler, Gabriel. "Driving Delhi? The Impact of Driving Restrictions on Driver Behavior" *Working Paper, in process*.

Kremer, M., & Miguel, E. (2007). The Illusion of Sustainability. *The Quarterly Journal of Economics*, *122*(3), 1007-1065.

Kremer, Michael, and Rachel Glennerster. "Improving Health in Developing Countries." *Handbook of Health Economics* 2 (2011): 201-315.

Kremer, Michael. "Randomized evaluations of educational programs in developing countries: Some lessons." *The American Economic Review* 93.2 (2003): 102-106.

Lucas, Adrienne M. and Isaac M. Mbiti. 2014. "Effects of School Quality on Student Achievement: Discontinuity Evidence from Kenya." *American Economic Journal: Applied Economics*, 6(3): 234-63.

Maluccio, John, and Rafael Flores. 2005. "Impact Evaluation of a Conditional Cash Transfer Program the Nicaraguan Red de Protección Social." Washington, DC: International Food Policy Research Institute (IFPRI) Report 141.

María Laura, Guillermo Cruces, and Laura Ripani. 2013. "Welfare programs and labor supply in developing countries: experimental evidence from Latin America." *Journal of Population Economics* 26(4): 1255-1284.

Mansilla, R., Köppen, E., Cocho, G., & Miramontes, P. (2007). On the behavior of journal impact factor rank-order distribution. *Journal of Informetrics*, *1*(2), 155-160

McKenzie D., S. de Mel, and C. Woodruff (February 2012). "One-time transfers of cash or capital have long-lasting effects on microenterprises," *Science*, 335(24): 962-66.

Meager, Rachael. "Understanding the impact of microcredit expansions: A Bayesian hierarchical analysis of 7 randomised experiments." *Working Paper* (2016).

Muralidharan, Karthik and Venkatesh Sundararaman. "The aggregate effect of school choice: Evidence from a two-stage experiment in India." *Quarterly Journal of Economics 130.3* (2015)*: 1011-66*

Muralidharan, Karthik, Paul Niehaus, and Sandip Sukhtankar. "Building State Capacity: Evidence from Biometric Smartcards in India" *American Economic Review, forthcoming* (2016).

Nadel, Sara, and Lant Pritchett. "Learning about Program Design With Rugged Fitness Spaces." *Unpublished,* 2016.

Neyman, Jerzy. 1923 [1990]. "On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9." *Statistical Science* 5 (4): 465–472. Trans. Dorota M. Dabrowska and Terence P. Speed.

Olken, Benjamin. A. "Monitoring Corruption: Evidence from a Field Experiment in Indonesia" *Journal of Political Economy*, vol. 115, issue 2, April 2007, pp. 200-249

Olken, Benjamin. A., Adnan Khan and Asim Khwaja. "Tax Farming Redux: Experimental Evidence on Performance Pay for Tax Collectors" *Quarterly Journal of Economics*, volume 131, issue 1, February 2016, pp 219-271.

Padró-i-Miquel, Gerard, Nancy Qian, and Yang Yao. "Social Fragmentation, Public Goods and Elections: Evidence from China.", Working Paper (2014).

Patel, Vikram, Helen A Weiss, Neerja Chowdhary, Smita Naik, Sulochana Pednekar, Sudipto Chatterjee, Mary J De Silva, Bhargav Bhat, Ricardo Araya, and Michael King. 2010. "Effectiveness of an Intervention Led by Lay Health Counsellors for Depressive and Anxiety Disorders in Primary Care in Goa, India (MANAS): A Cluster Randomised Controlled Trial." *The Lancet* 376 (9758): 2086–95.

Pritchett, Lant. "It pays to be ignorant: a simple political economy of rigorous program evaluation." *The Journal of Policy Reform* 5, no. 4 (2002): 251-269.

Rao, Gautam, Frank Schilbach, and Heather Schofield. "Sleepless in Chennai: the Economic Effects of Sleep Deprivation among the Poor." *Working paper, in process*.

Ravallion, M. (2012). Fighting poverty one experiment at a time: A review of Abhijit Banerjee and Esther Duflo's poor economics: A radical rethinking of the way to fight global poverty. *Journal of Economic Literature*, *50*(1), 103-114.

Rigol, Natalia, Reshmaan Hussam and Giovanni Regianni. "Slipped My Mind: Handwashing and Habit Formation." *Working paper, in progress.*

Rubin, Donald B. (1981). "Estimation in parallel randomized experiments. *Journal of Educational and Behavioral Statistics*, 6(4), 377-401.

Rubin, Donald B. "Estimating causal effects of treatments in randomized and nonrandomized studies." *Journal of educational Psychology* 66, no. 5 (1974): 688.

Schilbach, Frank. "Alcohol and self-control: A field experiment in India." *Working Paper* (2015).

Schultz, T. Paul. "School subsidies for the poor: evaluating the Mexican Progresa poverty program." *Journal of development Economics* 74.1 (2004): 199-250.

Shah, Neil Buddy, Paul Wang, Andrew Fraker, and Daniel Gastfriend. 2015. "Evaluations with impact Decision-focused impact evaluation as a practical policymaking tool." *International Initiative for Impact Evaluation Working Paper 25*.

Sim, Ida, et al. "Clinical decision support systems for the practice of evidence-based medicine." Journal of the American Medical Informatics Association 8.6 (2001): 527-534. APA

Todd, Petra E. and Kenneth I. Wolpin. 2006. "Assessing the Impact of a School Subsidy Program in Mexico: Using a Social Experiment to Validate a Dynamic Behavioral Model of Child Schooling and Fertility." *American Economic Review*, 96(5): 1384-1417.DOI: 10.1257/aer.96.5.1384

Udry, Christopher. "Risk and saving in Northern Nigeria." *The American Economic Review* 85.5 (1995): 1287-1300.

Vickrey, William (1969). Congestion Theory and Transport Investment. The American Economic Review, 59(2): 251-260.

Vivalt, E. (2015). How Much Can We Generalize from Impact Evaluations? Are They Worthwhile?. *Unpublished Manuscript*.

"World Malaria Report 2015." *World Health Organization*. WHO, Dec. 2015. Web. 31 May 2016.
<http://who.int/malaria/publications/world-malaria-report-2015/report/en/>.

Figures & Tables:

Number of Published RCTs

Evaluations by Type

2.4%

16.1%

16.7%

66.4%

8.3%

- ■ Differences in Differences
- ■ Instrumental Variables
- ■ RCT
- ■ Regression Discontinuity
- ■ Propensity Score Matching, or Other Matching Method
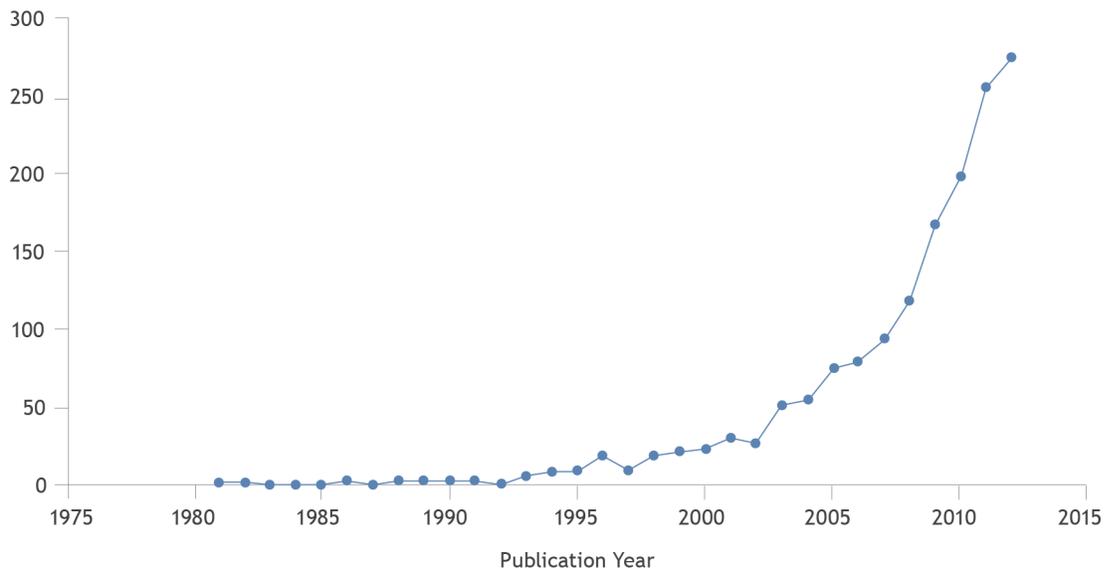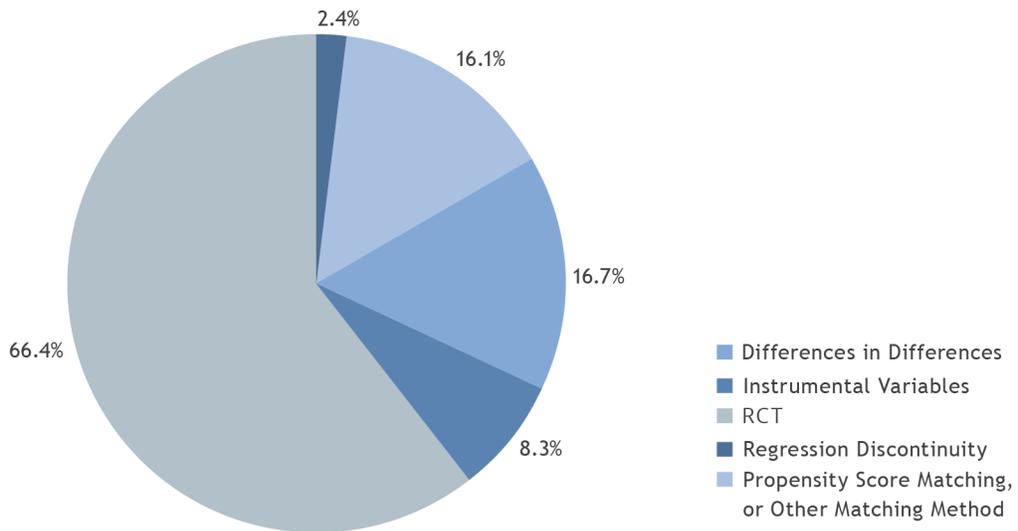
Figure 3B. Aidgrade.org Evaluations

Legend: Total Number of Evaluations · Total # of Econ Evaluations · Total # of Other Evaluations
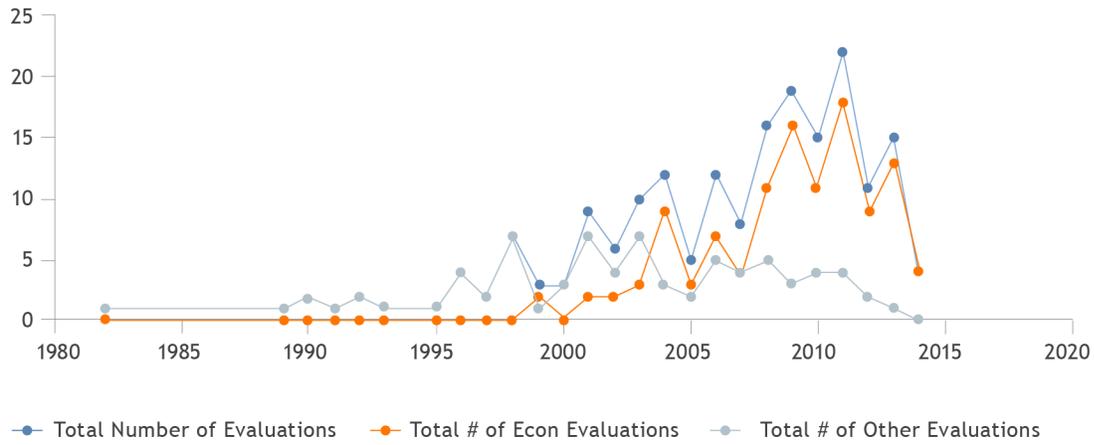
Figure 3A. Aidgrade.org Evaluations by Type

Legend: Percent RCTs in Econ · Percent RCTs in Other Fields · Non-RCTs
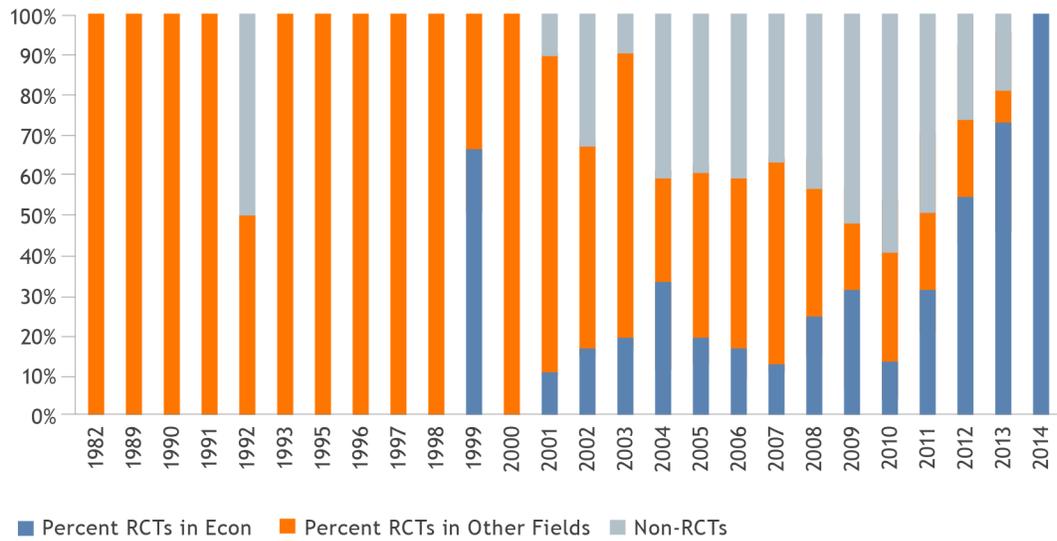
Figure 4. Fraction of BREAD Affiliates & Fellows with 1 or more RCTs

PhD Year
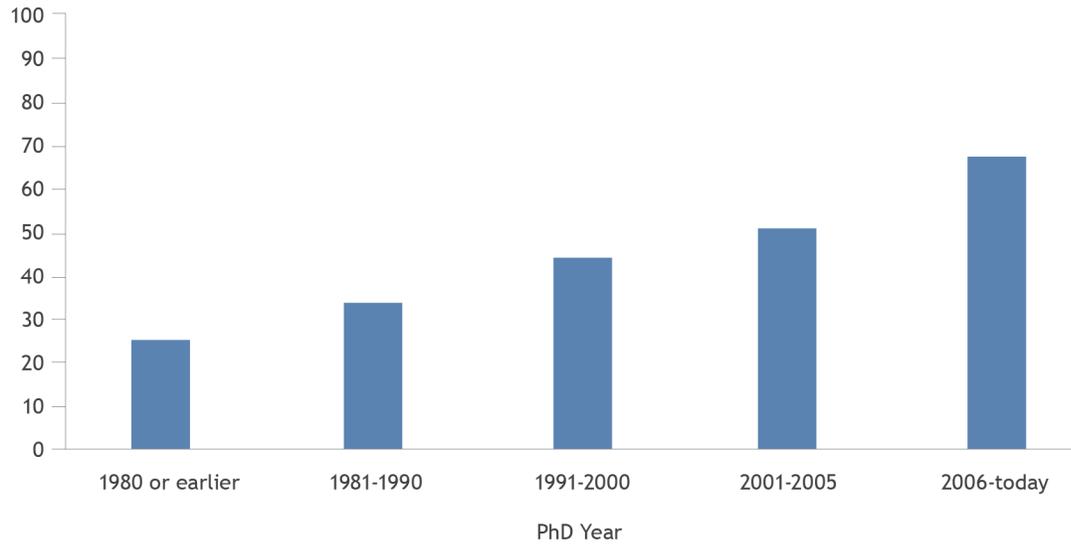
* Total Number of Fellows and Affiliates is 165

Figure 5. Percent of BREAD Conference Papers using a RCT
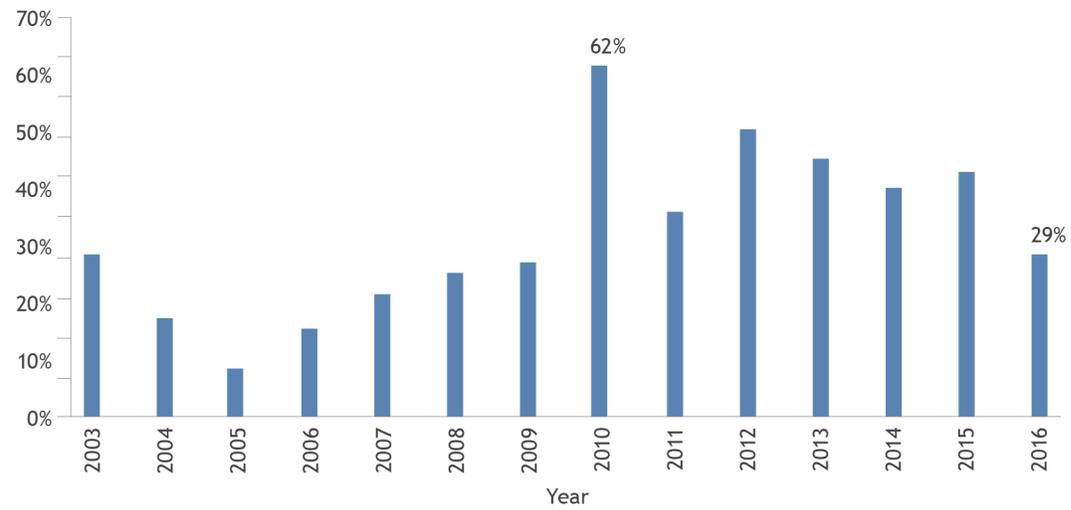
Year

Table 1: Papers in Top Journals

Journal	Year	Total # of Papers	# of Development Papers	# of which are RCTs
AER	2015	101	15	4
	2000	48	6	0
	1990	57	2	0
QJE	2015	40	1	1
	2000	43	5	0
	1990	52	3	0
JPE	2015	36	4	3
	2000	51	7	0
	1990	65	9	0
Restud	2015	48	7	2
	2000	36	3	0
	1990	40	1	0
Econometrica	2015	46	5	0
	2000	37	0	0
	1990	64	2	0
Total	2015	271	32	10
	2000	215	21	0
	1990	278	17	0

Table 2: NEUDC Conference Papers

Year	Total # of RCTs	Percent RCTs
2015	40	18.2%
2014	36	17.9%
2013	49	24.3%
2012	27	16.0%

Table 3: Future reach of DIV projects, by award type

Award Stage	Number of Awards	Total Awarded Value	Fraction Reaching More than 100,000 people	Fraction Reaching More than 1,000,000 people	People Reached[12]	DIV Expenditure per Person Reached
Stage 1 (< $100,000)	23	$2,353,136	17% (4/24)	8% (2/24)	6,723,733	$0.35
Stage 2 (<$1,000,000)	19	$9,557,926	44% (8/18)	11% (2/18)	16,931,044	$0.56
Stage 3 (<$15M)	1	$5,516,606	100% (1/1)	100% (1/1)	1,750,000	$3.15

---

[12] Two innovations (Voter Information Report Cards and CommCare) that reached over 100,000 people received both a Stage 1 and a Stage 2 award. In both of these cases, people reached by those innovations are counted as people reached by Stage 2 awards.

Appendix One: Data on All 43 DIV Awards 2010-2012

The table below shows all 43 DIV awards through December 31, 2012 and provides information on the number of people reached by each innovation that reached over 100,000 people. As discussed in the previous section, this includes both those reached through the original innovation and those reached through adapted forms of the innovation. The unit of analysis is the award; in two cases, DIV made two awards for the same innovation during this period.[1]

---

[1] This means that the overall impact cannot be estimated by summing the reach for each award.

DIV AWARDS: 43 AWARDS FROM 2010-2012

Project Name	Grantee	Country	DIV Stage	DIV Award Start Date	DIV Award Value	Development Economics Researchers Involved[2]	Reach if over 100,000	Key Entities Involved in Scaling	Type of Organization
The E-Bike: Practical, Scalable, Pollution-Free Mobile Transportation	SiGNa Chemistry, Inc.	U.S.	1	10/1/2010	$100,000	No			For profit
Cellular Monitoring: Improving Governance in Afghanistan	The Regents of the University of California, San Diego	Afghanistan; the innovation later expanded to Kenya, Uganda, South Africa	1	9/30/2010	$99,992	Yes	6,483,633[3]	Qualcomm, Inc.; HTC	Academic
	Dimagi Inc.	India	1	11/1/2010	$99,624	Yes	649,896		For profit

---

[2] This column is marked 'Yes' if development economics researchers were involved at any stage in the project development, implementation, or evaluation. In nearly all cases, development economics researchers were involved in conducting an RCT of the innovation.

[3] People reached include those reached under the DIV-supported innovation, and those reached under adaptations of the innovation which were used in other countries and in Afghanistan during a later election.

DIV AWARDS: 43 AWARDS FROM 2010-2012

Project Name	Grantee	Country	DIV Stage	DIV Award Start Date	DIV Award Value	Development Economics Researchers Involved[2]	Reach if over 100,000	Key Entities Involved in Scaling	Type of Organization
Scaling CommCare for Community Health Workers in India			2	8/31/2012	$996,424	Yes		Gates Foundation; Government of Bihar; BBC World Service Trust; Catholic Relief Services; IntraHealth	For profit
Improving Governance and Public Service Delivery with Voter Information Campaigns	Institute for Financial Management and Research - IFMR	India	1	9/30/2010	$98,957	Yes	10,253,704	Satark Nagrik Sanganathan (SNS); Hindustan newspaper	NGO
			2	2/7/2012	$200,000	Yes			NGO
Rural Solar Accessibility via Consumer Cooperative Enhanced Society (ACCESS) Retails	Lighting Rural Uganda with Solar	Uganda	1	10/4/2010	$98,360	No			NGO

DIV AWARDS: 43 AWARDS FROM 2010-2012

Project Name	Grantee	Country	DIV Stage	DIV Award Start Date	DIV Award Value	Development Economics Researchers Involved[2]	Reach if over 100,000	Key Entities Involved in Scaling	Type of Organization
Inventory Credit: Combining Storage and Savings To Increase Income	Innovations for Poverty Action	Sierra Leone	2	9/30/2010	$230,145	Yes			NGO
Proteinuria Self-Test For Early Detection of Pre-Eclampsia	Jhpiego Corporation	Nepal	1	1/15/2011	$100,000	No			NGO
Examining barriers to fertilizer use in Kenya	Innovations for Poverty Action	Kenya	1	9/7/2011	$99,828	Yes			NGO
Developing an Affordable Balloon Tamponade for Postpartum Hemorrhage Treatment and Management	Program for Appropriate Technology in Health (PATH)	Ghana	1	10/13/2011	$99,793	No			NGO

DIV AWARDS: 43 AWARDS FROM 2010-2012

Project Name	Grantee	Country	DIV Stage	DIV Award Start Date	DIV Award Value	Development Economics Researchers Involved[2]	Reach if over 100,000	Key Entities Involved in Scaling	Type of Organization
Developing Sustainable Sanitation in Urban Slums	Sanergy	Kenya	1	11/17/2011	$99,840	No			NGO
Life changing and Revenue generating Electricity for Sub-Saharan Africa: EGG-energy's Franchised Solar Hubs	EGG-energy	Tanzania	1	2/21/2012	$100,000	No			For profit
Increased Uptake and the Use of Safe Water Filters at Scale	RAND Corporation	Kenya	1	8/29/2012	$108,735	Yes			NGO
Digital Attendance and Medical Information System in PHCs	Institute for Financial Management and Research - IFMR	India	2	9/30/2010	$172,679	Yes	1,800,000	Government of Karnataka, National Rural Health Mission of Karnataka	NGO

DIV AWARDS: 43 AWARDS FROM 2010-2012

Project Name	Grantee	Country	DIV Stage	DIV Award Start Date	DIV Award Value	Development Economics Researchers Involved[2]	Reach if over 100,000	Key Entities Involved in Scaling	Type of Organization
Leveraging Public-Private Partnerships for the Indian and Global Environment	Institute for Financial Management and Research - IFMR	India	2	1/27/2012	$185,553	Yes			NGO
Viability of Cyanobacterial Bio-fertilizer to Improve Soil Fertility and Crop Yields in Ethiopia	Thin Air Nitrogen Solutions, LLC	Ethiopia	1	3/1/2012	$99,854	No			For profit
Household Hand-Washing Device - Commercial Development	WaterSHED	Vietnam	1	4/24/2012	$100,000	No			NGO
Renewable Powered Micro Grids for Rural Lighting	Mera Gao Micro Grid Power (MGP) Private Limited	India	2	9/30/2011	$300,000	No	120,000	Mera Gao Power	For profit

DIV AWARDS: 43 AWARDS FROM 2010-2012

Project Name	Grantee	Country	DIV Stage	DIV Award Start Date	DIV Award Value	Development Economics Researchers Involved[2]	Reach if over 100,000	Key Entities Involved in Scaling	Type of Organization
Improving patient safety in Pakistan's hospitals: a mobile health solution for enabling and monitoring the 'First, do no harm' principle	Indus Hospital	Pakistan	1	7/2/2012	$99,250	No			NGO
Smoothing the Costs of Education: Microsavings in Ugandan Primary Schools	Innovations for Poverty Action	Uganda	2	3/27/2012	$181,537	Yes			NGO
InSight: Mobile Accounting and Financial Inclusion in Emerging Markets	InVenture	India	1	9/30/2012	$100,000	No			For profit
Milele Tube Final Testing and Marketing Introduction	Baisikeli Ugunduzi	Kenya	1	10/1/2012	$100,000	No			NGO

DIV AWARDS: 43 AWARDS FROM 2010-2012

Project Name	Grantee	Country	DIV Stage	DIV Award Start Date	DIV Award Value	Development Economics Researchers Involved[2]	Reach if over 100,000	Key Entities Involved in Scaling	Type of Organization
Testing a Digital Platform's Ability to Recreate the Success of Rural Community-Led Total Sanitation in Urban Communities	IDEO.org	Ghana	1	9/12/2012	$100,000	No			NGO
Developing a Supply Chain for Hermetic Storage of Grain in Afghanistan	Purdue University	Afghanistan	1	9/9/2011	$88,400	No			Academic
The Entrepreneurial Finance Lab Research Initiative	President and Fellows of Harvard College	Originally Egypt, Later Reallocated	2	4/24/2012	$438,002	Yes			Academic
Consumer Action and Matatu Safety	Georgetown University	Kenya	2	8/25/2011	$291,154	Yes	3,000,000	Insurance companies; Safaricom; Government of Kenya	Academic

DIV AWARDS: 43 AWARDS FROM 2010-2012									
Project Name	Grantee	Country	DIV Stage	DIV Award Start Date	DIV Award Value	Development Economics Researchers Involved[2]	Reach if over 100,000	Key Entities Involved in Scaling	Type of Organization
Remittances for Educational Finance	University of Michigan	Philippines	1	2/17/2012	$96,409	Yes			Academic
Mobile Phone Agriculture Extension: Using ICT to Reduce Outreach and Monitoring Costs	Innovations for Poverty Action	Kenya	1	9/15/2011	$96,394	Yes			NGO
Fighting Tuberculosis through Community Based Counselors in Northern Indian slums: A Randomized Evaluation of Performance Based Incentives	Institute for Financial Management and Research - IFMR	India	1	10/1/2012	$75,104	Yes			NGO

DIV AWARDS: 43 AWARDS FROM 2010-2012									
Project Name	Grantee	Country	DIV Stage	DIV Award Start Date	DIV Award Value	Development Economics Researchers Involved[2]	Reach if over 100,000	Key Entities Involved in Scaling	Type of Organization
The Role of Mobile Banking in Expanding Trade Credit & Business Development in Kenya	Innovations for Poverty Action	Kenya	2	8/24/2011	$360,195	Yes			NGO
Recruiting and Compensating Community Health Workers: A National Field Experiment in Zambia	Innovations for Poverty Action	Zambia	1	9/5/2012	$99,032	Yes	240,100	Zambia Ministry of Health	NGO
Affordable Access to Energy for All: Innovative Financing for Solar Systems	d.light design	Uganda	2	9/20/2012	$1,020,126	Yes	825,000		For profit
Evaluating the Impact of Mobile Banking and Business Skills on Microenterprise Development	Faculdade de Economia da Universidade Nova de Lisboa	Mozambique	2	9/30/2012	$293,146	Yes			Academic

DIV AWARDS: 43 AWARDS FROM 2010-2012

Project Name	Grantee	Country	DIV Stage	DIV Award Start Date	DIV Award Value	Development Economics Researchers Involved[2]	Reach if over 100,000	Key Entities Involved in Scaling	Type of Organization
Turning the Tap Off on Multi Drug Resistant TB and Extensively Drug Resistant TB	Operation ASHA	India/Cambodia	2	10/1/2012	$897,324	Yes			NGO
Improving Health Service Delivery Through Community Monitoring and Non-Financial Awards	Innovations for Poverty Action	Sierra Leone	2	10/1/2012	$432,258	No			NGO
Scaling Biochar: Investing in soils, improving livelihoods and sequestering carbon	Innovations for Poverty Action	Kenya	1	9/20/2012	$99,952	Yes			NGO
Developing and Testing a Sustainable Distribution Model for Improved Cookstoves in Darfur and Ethiopia	Potential Energy	Ethiopia, Sudan	2	4/18/2012	$1,500,000	Yes	194,542		NGO

DIV AWARDS: 43 AWARDS FROM 2010-2012									
Project Name	Grantee	Country	DIV Stage	DIV Award Start Date	DIV Award Value	Development Economics Researchers Involved[2]	Reach if over 100,000	Key Entities Involved in Scaling	Type of Organization
Women's Network to Improve Clean Energy	Solar Sister, Inc.	Uganda South Sudan Tanzania	2	5/28/2012	$1,000,000	No			NGO
Chlorine Dispensers: Bringing Safe Water to Scale	Innovations for Poverty Action (Evidence Action)	Kenya, Uganda, Rwanda, others TBD (Malawi later substituted for Rwanda)	3	6/20/2012	$5,516,606	Yes	2,610,000	Evidence Action; Abt Associates; Liberty Foundation; Skoll Foundation; MercyCorps; Stone Family Foundation; Ceniarth Foundation; Kiva; One Acre Fund; Gov't of Malawi (Ministry of Health); Gov't of Kenya	NGO

DIV AWARDS: 43 AWARDS FROM 2010-2012									
Project Name	Grantee	Country	DIV Stage	DIV Award Start Date	DIV Award Value	Development Economics Researchers Involved[2]	Reach if over 100,000	Key Entities Involved in Scaling	Type of Organization
Scaling BoPtical Care Hub Model in India	VisionSpring	India	2	8/20/2012	$585,350	No	274,837		NGO
How to Make Better Businesses: Experimental Evidence of the Components of Entrepreneurship	Innovations for Poverty Action	Uganda	1	8/27/2012	$93,612	Yes			NGO
Ghana National Apprenticeship Program Impact Evaluation: Effort, Incentives and Returns	Innovations for Poverty Action	Ghana	2	10/1/2012	$474,033	Yes			NGO
Total			43		$17,427,668	26	24,110,313		

Appendix Two: Detailed Information on DIV-Supported Innovations That Have Reached More Than 100,000 People

The table below shows all 43 DIV awards through December 31, 2012 and provides information on the number of people reached by each innovation that reached over 100,000 people. As discussed in the previous section, this includes both those reached through the original innovation and those reached through adapted forms of the innovation. The unit of analysis is the award; in two cases, DIV made two awards for the same innovation during this period.[4]

---

[4] This means that the overall impact cannot be estimated by summing the reach for each award.

Improving Governance and Public Service Delivery with Voter Information Campaigns

http://www.usaid.gov/div/portfolio/right-vote-right (Stage 1) http://www.usaid.gov/div/portfolio/improving-governance-and (Stage 2)
DIV contribution: Stage 1 and 2 Funding
Award Amount: $98,957 (Stage 1), $200,000 (Stage 2)
Award Period: 2010-2012

Problem: Lack of voter information may lead to poor electoral outcomes, potentially contributing to corruption and other forms of poor governance.

Innovation: Researchers conducted a multi-year project to test 1) whether better electoral outcomes can be achieved by directly providing voters with information, either on politician responsibilities or on actual politician performance and qualifications, 2) whether anticipation of and actual public disclosures on responsibilities and/or performance can cause incumbents to improve their service delivery and performance and change decisions on whether to stand for re-election, and 3) whether governance can be strengthened by directly providing elected officials with information about the quality of service and if this, in turn, affects usage of these amenities.

Impact: An initial evaluation (not supported by DIV) showed that voters respond to information about elected officials: Banerjee, Pande, Kumar, and Su (2010) show that an information campaign increased voter turnout by 3.5 percent, or two percentage points (from 57.5% to 59.5%). While the campaign did not influence average incumbent vote share, worse performing incumbents and those facing better-qualified challengers received significantly fewer votes. Cash-based vote-buying was 19 percentage points less likely to occur in treatment polling "stations". Subsequently, DIV provided funding for a second evaluation designed to shed light on questions 2) and 3) in a Stage 1 grant, followed by Stage 2 funding for a follow-up evaluation to test the impact of voter education campaigns in Bihar, one of the poorest states in India. Official results for those projects are still pending and the researcher has requested that interim results not be provided in this document, although they have been provided to DIV and shared with others.

Reach: Satark Nagrik Sangathan (SNS), an Indian civil society organization that promotes transparency and accountability in government collaborated with J-PAL South Asia in the design of the research to create information report cards for councilor elections in Delhi and MP elections across the country. Newspapers have published this content, presumably because it was free and they believed it would be of interest to their readers. Based on the outlets where the information was sent and its circulation, we estimate that since 2012 (the first elections to take place after the first DIV award to

support this innovation), these efforts have reached **approximately 10 million voters[5].** Note, however, that this program was less intensive and costly than that examined under the original evaluation, because it did not involve field workers dropping off the newspapers door to door, and the impact may thus differ. SNS also circulated the reports cards via internet and television stations broadcast report card information, but these sources of reach are not included in the totals below.

Likely Future Reach and Sustainability: Prospects for sustainable future reach exist. Follow-up evaluations in Bihar are forthcoming. SNS's commitment to the voter report card program and newspapers' interest in publishing report cards suggests that similar interventions are likely in future elections.

Campaign	Newspaper	Circulation	Voting Age Adults per hh	Election Turnout	Estimated # of Voters Reached
Councillor Report Cards 2012	Hindustan Times Hindi (Delhi edition)	299,576	2	0.58	347,508
	Hindustan Times (Delhi edition)	964,500	2	0.58	1,118,820
MP Report Cards 2014	India Today magazine - Hindi (National Edition)	494,000	2	0.664	656,032
	Hindustan Times - Hindi (national edition)	6,123,000	2	0.664	8,131,344
Total					**10,253,704**

---

[5] Assumptions: 1. People reached is a function of total circulation of newspapers and magazines in which report cards were printed, multiplied by the number of voting-age adults per household, and adjusted by the turnout in that particular election. 2. In cases where readership but not circulation figures were available, circulation was estimated by dividing by 2. On average, we calculate readership to be about 2 x circulations, based on data on several Indian newspapers provided by Business Standard (http://www.business-standard.com/content/general_pdf/021012_01.pdf). 3. Average household size in India according to the 2011 census is 4.8. We assume that, on average, there are 2 voting age adults (18+) per household.

Collecting and Transmitting Polling Station Vote Counts: Improving Governance in Afghanistan [6]

http://www.usaid.gov/div/portfolio/caught-camera-testing
DIV contribution: Stage 1 Funding
Award Amount: $99,922
Award Period: 2010 – 2011

Problem: Irregularities plague elections in emerging democracies, undermining voters' ability to hold their leaders accountable.

Innovation: One low-cost alternative to having international election observers is to use mobile technology to record and transmit information about votes cast at specific polling stations. Researchers designed an anti-fraud technology called "photo quick count," which allows local election monitors to photograph provisional vote tally sheets at individual polling centers and compare them to the official vote count after aggregation. (In a clean election, the before and after tallies should be identical.) Letters announcing the photographic vote count verification were sent to a random sample of polling stations during the 2010 parliamentary elections in Afghanistan. This study covered 471 polling stations, about 5% of the national sample.

Impact: Results from a randomized impact evaluation show that announcing the photographic quick counts reduced the theft of election materials by 60% and reduced votes for politically powerful candidates by 25%. Researchers estimate that between 9.37 and 17.17 fewer votes changed during the aggregation process in cases where candidates had close ties to the official in charge of aggregating votes.

Reach: During the 2010 parliamentary elections in Afghanistan, 326,661 people[7] voted at polling stations where photographic monitoring took place. Ashraf Ghani's political campaign implemented a modified version of the "photo quick count" monitoring system, recording votes for the purposes of comparison using paper and clipboards rather than photographs, during the first round of the 2014 Presidential election in Afghanistan. Campaign staff indicate reaching as much as 80 percent of the polling centers operating on Election Day in Afghanistan, potentially having an impact on as many as 5.25 million voters.[8] Though Ashraf Ghani ultimately won the presidency, the subsequent runoff election was marred by charges of fraud on both sides. Photo quick count was not, to our knowledge, used in the runoff election. The use of campaign staff (versus local election monitors) and the use of paper and clipboards may mean the impact in this campaign was different than in the 2010 parliamentary elections.

---

[6] Long, James D. "Innovations to Reduce Electoral Corruption and Increase Citizen Engagement Using Information and Communications Technology and Digital Media in Emerging Democracies."
June 25, 2014.

[7] Email from James Long, August 13, 2014.

[8] Email from Michael Callen, December 8, 2014.

Building on the positive results from the 2010 trial in Afghanistan, the approach was replicated in Uganda during the 2011 Presidential and Parliamentary elections, in Kenya during the 2013 national election, and in South Africa during the 2014 national election. These efforts have reached over 2.8 million people since the original DIV-supported evaluation.

In Uganda, field teams were given smartphones equipped with a mobile app designed for collecting photo quick counts. They visited 1,001 polling stations (approximately 5% of the national sample), reaching 412,625 voters.

In Kenya, field teams visited 1,200 polling stations (approximately 4% of the national sample), reaching 494,347 voters.

In South Africa, **1,936,450 people voted at polling stations where smartphone monitoring technology was being used during the 2014 national election.** With Stage 2 funding from DIV, the use of information and communications technology (ICT) is being expanded to enable voters to engage in the electoral process in the lead-up to and during the 2014 national election in South Africa. A randomized evaluation is currently being conducted and results are forthcoming.

Likely Future Reach and Sustainability: Prospects for sustainable future reach exist. Election monitoring technology was used by current Afghan president Ashraf Ghani's political party during the 2014 Afghan presidential election and similar election monitoring technology has been used in three other countries with promising results. To our knowledge, however, no international election monitoring organization has adopted this approach.

Additional Resources:
J-PAL Evaluation Summary: http://www.povertyactionlab.org/evaluation/election-fraud-and-government-legitimacy-afghanistan.

Digital Attendance and Medical Information System in Primary Health Clinics

http://www.usaid.gov/div/portfolio/thumbs-monitoring-health
DIV contribution: Stage 1 Funding
Award Amount: $172,679
Award Period: 2011-2013
Partner: Government of Karnataka (Health Department)

Problem: The government is the primary health care provider in many developing countries. However, despite large budgetary allocations and an extensive network of public health centers, health outcomes often remain poor. Evidence suggests that high rates of staff absenteeism, enabled by poor systems of accountability, may be a reason why health care systems do not function as intended. India has a vast network of over 24,000 public Primary Health Centers (PHCs) that provide basic services such as immunization and antenatal care to the poor. Yet absenteeism is high in PHCs across the country. In the PHCs in this study, doctors were present only 36 percent of the time in 2010. As a result, many citizens coming to receive care leave unattended, and often end up visiting untrained or unqualified providers for treatment.

Innovation: The government of Karnataka state in India partnered with researchers to implement and evaluate a biometric monitoring system that objectively records attendance and reports it to supervisors in real time, combined with a robust system of incentives and penalties for unauthorized absences to improve staff attendance and patient health.

From a sample of 322 PHCs across five socio-economically diverse districts, 140 were randomly selected to receive the biometric devices consisting of a fingerprint reader and a mobile phone, while the remaining 182 continued with the status quo paper system of marking attendance. The device was used to record staff attendance via thumb impression at the beginning and end of each day. It was also capable of recording details about cash benefits paid to patients along with photographs and signatures and thumb impressions of beneficiaries taken at the clinic, and statistics regarding number of patients seen and the diseases treated. In practice it was primarily used for attendance monitoring.

Attendance data could be transferred wirelessly using the existing cellular network to the state health headquarters in Bangalore so supervisors could track staff attendance in near real time. This data was analyzed and processed and then communicated back to the districts. This attendance information was coupled with an extensive system of incentives and penalties to encourage better attendance. Based on the attendance data, the government planned to issue both positive incentives, such as awards for staff members with good attendance records, as well as negative incentives, such as reprimand letters, disciplinary action, suspension from service, docking of pay, and deduction of earned leave for employees with unauthorized absences.

Impact: Even though the official leave policy was not strictly enforced, the monitoring system increased medical staff attendance by 3.4 percentage points (8.7 percent) relative to medical staff in comparison PHCs. It had the greatest impact on nurses, lab technicians, and pharmacists (an 18 percent increase), but virtually no effect on doctors. The new monitoring system seemed to have led to improved antenatal care and infant health. Pregnant women in treatment PHCs were 10.6 percentage points (27 percent) more likely to receive recommended Iron Folic Acid tablets. Relative to expectant mothers in comparison PHCs, mothers in treatment PHCs were 8 percentage points (16 percent) more likely to have their baby delivered by a doctor. Their newborn children were 4.6 percentage points (26 percent) less likely to be born underweight and weighed 67 grams more on average. The main finding of the project is that even though the top-down monitoring is very hard to sustain, even small gains in attendance can have large impacts on health outcomes.

Reach: During the intervention (spanning about 18 months), the primary health centers in the treatment group attended to about 1,800,000 patients. Given the challenges of hiring and retaining doctors willing to work in remote health clinics and its inability to effectively monitor, the government of Karnataka state decided in 2013 not to scale up the digital attendance monitoring program that had been piloted. In 2014, Government of India started rolling out digital attendance gathering, but in the much easier to monitor context of staff working in government offices in cities like the national capital Delhi. It is not clear how effective the monitoring is, and whether it has led to an increase in attendance.

Likely Future Reach and Sustainability: Prospects for sustainable future reach are uncertain. Karnataka decided to not scale up this program because of difficulties in retaining doctors and in maintaining effective monitoring. The government of India has implemented a similar program to monitor government offices in larger cities. More time is needed to determine if attendance monitoring in different contexts will be effective and sustainable.

Scaling CommCare for Community Health Workers in India

http://www.usaid.gov/div/portfolio/scaling-commcare-deliver
DIV contribution: Stage 1 and 2 Funding
Award Amount: $99,624 (Stage 1), $996,424 (Stage 2)
Award Period: 2010 – ongoing

Problem: In India and elsewhere, resource-strapped public health departments and supporting NGOs are increasingly reliant on community health workers (CHWs) to provide life-saving services in areas that traditional medical providers struggle to reach. However, training, supervision, and evaluation of CHWs are major challenges to ensuring that people served by CHWs receive the same level of care as those able to access treatment from traditional medical providers. Mobile-based solutions offer a potentially cost-effective and scalable way to empower CHWs to deliver better care.

Innovation: CommCare is a mobile platform that enables CHWs to enroll and manage clients, to create patient intake forms, to conduct more timely visits, and to access learning resources with information about healthy behavior. Developed by Dimagi, a social enterprise that makes open source software to improve healthcare in developing countries and for the underserved, CommCare provides actionable data to help CHWs improve their performance. CHWs can submit patient data in real-time to a central cloud server, where it is privacy-protected and backed up. Supervisors can view each CHW's performance indicators, including daily activity, number of clients, length of visits, and follow-up rates.

Impact: A randomized controlled trial showed that reminders from CommCare, with eventual escalation to supervisor notification, led to 85% more timely visits compared to CHWs not receiving reminders.[9] Additional studies have shown that CommCare helps increase CHWs' knowledge of health risks and improve data completeness and accuracy. Dimagi is continually evaluating, publishing, and presenting results and lessons learned from deploying and scaling CommCare, and CommCare has more evidence attesting to its effectiveness than any other mHealth platform.[10]

Stage 2 expansion is being accompanied by several ongoing randomized evaluations in partnership with CARE and Mathematica Policy Research.

---

[9] B. DeRenzi, L. Findlater, G. Borriello, J. Jackson, J. Payne, B. Birnbaum, T. Parikh, N. Lesh, "Improving Community Health Worker Performance Through Automated SMS", ICTD 2013.
[10] Chatfield, A., Javetski, G., Lesh, N., CommCare Evidence Base. February 2013.
  https://confluence.dimagi.com/download/attachments/12226140/CommCare+Evidence+Base+February+2013.pdf?version=1&modificationDate=1363619871232

Reach: In 2010, Dimagi received Stage 1 funding from DIV to establish proof of concept for CommCare. Dimagi was able to support 11 organizations in 8 states in India. Six of these projects from the Stage 1 grant continue to be active, and 500 CHWs used CommCare to reach over 32,000 patients. After a successful initial grant, Dimagi received Stage 2 funding to increase their field team, further develop the evidence base for CommCare, and build the organizational capacity to expand CommCare across India. As of May 2014, Dimagi introduced CommCare to 58 organizations within India, reaching 2,736 CHWs and 530,340 patients.[11]

Likely Future Reach and Sustainability: Prospects for sustainable future reach exist. Increased smartphone usage will make it easier for community health workers to adopt mobile phone software services. Significant demand from large institutional actors may exist for this product (for example, several USAID offices, overseas missions, and implementing partners are highly interested in procuring and implementing CommCare for use in a range of projects).

---

[11] USAID DIV Milestone 12 – PMP – 14 July 2014. Obtained from Dimagi.

Affordable Access to Energy for All: Innovative Financing for Solar Systems

https://www.usaid.gov/div/portfolio/affordable-access-energy
DIV contribution: Stage 2 funding
Award Amount: $1,020,126
Award Period: 2012-2014

Problem: In Africa, millions of people lack electricity.. People living "off-grid" typically rely on kerosene lanterns or diesel power generators, but such power sources can be detrimental in the long term, in both health and environmental impacts. However, healthier and safer alternatives to kerosene and diesel, such as solar power, remain out of reach for many because of their high upfront costs.

Innovation: d.light provides a solar power alternative to households throughout Uganda by offering its D20 solar technology with user-friendly financing. Its home solar system includes a solar panel, two fixed LED lights, a portable LED lantern and a mobile phone charger. Via flexible, pay-as-you-go financing, d.light's consumers avoid heavy up-front cost. d.light's financing mechanism is similar to a layaway model, with the added bonus of being able to use the product while paying it off.

Impact: An IDInsight study using a difference-in-difference methodology estimated d.light that consumers save $103 in household energy costs over a 5-year product lifetime.[12] . d.light estimates that switching to a solar home system results in annual CO_2 mitigation of 276kg per household. ID Insight reports that d.light customers report an 88.1 percent decrease in burns (6.39 percentage points) and a 93.3% decrease in household fires (5.96 percentage points), and that they consume 6.2 more hours of high quality light and 4.3 fewer hours of low-quality light, with a net increase of 2.9 hours of light. While there was no statistically significant impact on self-reported personal health, coughing decreased by 8.5 percentage points (p=0.052).

Reach: d.light has sold over 150,000 solar home systems, reaching approximately 825,000 individuals across Uganda and Kenya.

Likely Future Reach and Sustainability: Prospects for future reach exist. d.light is one of the largest solar product distributors, with products in 62 countries. d.light operates from ten field offices and multiple international hubs, with continued plans for expansion. d.light's expansion and new D30 system was recently funded with a DIV Stage 2 award for $1,000,000.

---

[12] One potentially important caveat is that ID Insight estimates household energy expenditures of $0.96 per week, whereas an analysis of data from the World Bank's Living Standards Measurement Survey (LSMS), which measures consumption of a wide range of goods, finds expenditures of only $0.41 per week, raising the possibility that ID Insight's estimates might have been subject to social desirability bias from respondents aware that the survey was designed to evaluate d.light's product. Of course d.light customers are less likely to be living in poverty than the average Ugandan, and self-selected into purchasing the product, so likely would spend more on energy than the average Ugandan in the absence of the product. Additionally, the LSMS does not measure household phone charging expenditures, which represent a significant contributor to ID Insight's estimate of $0.96 per week in energy expenditure.

Consumer Action and Matatu Safety[13]

http://www.usaid.gov/div/matatus
DIV contribution: Stage 2 and 3 Funding
Award Amount: $291,154 (Stage 2), **$2,990,570** (Stage 3)
Award Period: 2011-2013, 2014 – ongoing

Problem: Road traffic accidents are a leading cause of death among people aged 5-44 in Africa, and it is estimated that accidents cost the continent $10 billion per year. Many interventions to reduce road accidents have been undertaken in developed economies, including programs to reduce the volume of driving, to improve the safety features of road networks, and to enforce traffic regulations more effectively, but few studies have explored how such interventions might work in developing countries.

Innovation: Researchers partnered with a local NGO and Safaricom, a major telecom company, to design and implement a road safety messaging campaign in Kenya. "Speak Up!" stickers encouraging passengers to speak up against bad driving were placed in a random sample of minibuses, and drivers were rewarded through a lottery for keeping the stickers in place. These rewards ranged from US $25 to $60. The stickers, about 11 by 3 inches, were placed on the metal panel between a passenger window and the ceiling of the vehicle, ensuring that at least one sticker was within eyesight of each passenger sitting in the main cabin. The first study (prior to DIV funding) covered 2,400 matatus[14] operating along a set of long-distance routes.

Impact: The initial campaign was highly effective: using data from four large Kenyan insurance companies, compared to rates for buses without stickers, road accident insurance claims fell by over 50%, and claims involving injury or death dropped by 60%. The cost of the intervention was just under $2 per vehicle for the stickers, and $5 per vehicle per year for the lottery, or a total of $7,000 per 1,000 vehicles per year. Researchers estimated that the intervention saved 1200 disability-adjusted years of life (DALYs) at a cost of $5.80 per year. The Stage 2 intervention (the first funded by DIV) covered 12,000 vehicles from one large Kenyan insurance company. In this initial scale-up, impact dropped but was still substantial: compared to rates for buses without stickers, road accident insurance claims fell by over 25%. Researchers estimated the cost-effectiveness of the sticker intervention between $40 and $235 per DALY saved.

---

[13] Habyarimana, James and William Jack. 2011. "Heckle and Chide: Results from a randomized road safety intervention in Kenya." *Journal of Public Economics*. 95(11-12): 1438-1446.
  Habyarimana, James and William Jack. 2015. "Results of a large-scale randomized behavior change intervention on road safety in Kenya" PNAS 2015 112 (34) E4661-E4670.

[14] In Kenya, minibuses, called matatus, are a popular form of transportation. A fleet of roughly 30,000 matatus carries approximately 3 million passengers a day (100 each, 14 at a time).

Reach: In Stage 2, 800,000 people rode in matatus with "Speak Up!" stickers each day. An estimated 3 million commuters in Kenya rode in a matatu with a "Speak Up!" sticker over the life of the program.[15] The current intervention will be implemented in 35,000 vehicles in Kenya (20,000 with Direct Line Assurance and 15,000 through the National Transport and Safety Authority (NTSA)) and will be tested for expansion in Tanzania, Uganda and Rwanda. A formal Memorandum of Understanding (MoU) has been signed with Kenya's largest insurance provider (which has more than 50% of the market) to implement the program.

Likely Future Reach and Sustainability: Prospects for future reach exist. In May 2015, Georgetown University (gui²de) launched the Zusha! Road Safety Campaign[16] and intends to work with the government and civil society actors to place stickers in every matatu in the country. Kenya's largest insurance company distributes stickers as part of matatu coverage, and the government now installs stickers during annual vehicle inspections. DIV supported the Zusha! campaign with a $2,990,570 Stage 2 award in 2014. Full implementation will require continued involvement from insurers and/or the NTSA beyond the scope of this award phase. The project evaluation team has begun initial roll-out of an RCT testing the intervention in Tanzania and is in talks with potential adopters in Uganda and Rwanda.

---

[15] Email from Billy Jack, August 10, 2014.

[16] See http://gui2de.georgetown.edu/projects/zusha

Chlorine Dispensers: Bringing Safe Water to Scale [17]

http://www.usaid.gov/div/portfolio/chlorine
DIV contribution: Stage 3 Funding
Award Amount: $7,416,557
Award Period: June 2012 – January 2016

Problem: Diarrheal disease, which spreads through contaminated drinking water, kills nearly 1 million children under the age of 5 every year. Protecting communal water sources, for example by encasing them in concrete, is one way to prevent contamination. However, clean water stored in homes can become re-contaminated with a dirty cup or an unwashed hand. In response, the World Health Organization recommends the use of dilute chlorine, which not only disinfects water, but provides ongoing protection from recontamination for over 24 hours.

The standard approach to encourage rural populations to adopt chlorination has typically been via social marketing—the promotion and sale of small bottles of chlorine through the private sector. Despite the significant health benefits, and relatively low price of chlorine ($0.30 for a family of five for a month in Kenya[18]), adoption remains low.

Innovation: A free, point-of-collection water chlorination system was designed to address the issue of recontamination and low usage rates of dilute chlorine available for purchase. Chlorine dispensers are placed at water sources, which serve as a visual reminder to treat water when it is most salient—at the time of collection. The source-based approach makes drinking water treatment convenient because the dispenser valve delivers an accurate dose of chlorine to treat the most commonly used water collection containers, while the public nature of the dispenser also contributes to learning and habit formation. In addition, local promoters provide frequent reminders and encouragement to other community members to use the product. At scale, chlorine dispensers could cost less than $0.50 per person annually, making them one of the most cost effective ways to reduce diarrheal disease and save lives.

Impact: Results from a randomized impact evaluation conducted in Kenya (prior to DIV funding) showed that three to six months after introducing the point-of-collection chlorine dispensers, take-up was 60% compared to only 7% for socially marketed in-home chlorine treatment.[19] Several years

---

[17] Ahuja, Amrita, Michael Kremer, and Alix Peterson Zwane. 2010.Providing Safe Water: Evidence from Randomized Evaluations, *Annual Review of Resource Economics* 2: 237-256
[18] Kremer, Miguel, Null, and Zwane. "Water Technologies case study: what works best in poor countries," Boston Review, September/October 2008. Available online at
  http://bostonreview.net/BR33.5/miguel.php
[19] Ahuja et al., 2010.

later (following DIV funding), program monitoring of the at-scale dispensers program finds that take-up of chlorine averages 50% in Kenya, 47% in Uganda and 84% in Malawi.

In early 2016, when there were 2.4 million active chlorine users, Evidence Action estimated that their 27,000 point-of-collection dispensers in Kenya, Uganda, and Malawi avert nearly 570,000 cases of diarrhea and over 14,000 disability adjusted life years (DALYs), annually.

	Kenya	Uganda	Malawi	Total
# Dispensers installed	17,784	5,585	3,333	26,702
# Cases diarrhea averted	563,716	221,549	51,941	837,206
# DALYs averted	15,054	5,882	1,388	22,324
# People who use chlorine	1,031,380	813,740	520,040	2,365,160

Reach: The expansion of point-of-collection chlorine dispensers began in Kenya in 2008 and has expanded both within Kenya and also to Uganda and Malawi with the support of DIV. Approximately 4.5 million people now have access to water treatment solution from dispensers and 2.6 million consume treated water.

Likely Future Reach and Sustainability: Evidence Action, an organization *dedicated to bringing cost-effective and evidence-based interventions to scale, is spearheading these efforts and has already secured over $12 million in funding since its inception. Evidence Action has also pre-sold more than $13 million in carbon credits. Additional funding beyond carbon credits from governments or donors will likely be needed to sustain the approach.*

Additional Resources:
J-PAL Evaluation Summary: http://www.povertyactionlab.org/evaluation/source-dispensers-and-home-delivery-chlorine-kenya
J-PAL Policy Bulletin: http://www.povertyactionlab.org/publication/the-price-is-wrong

Recruiting Community Health Workers in Zambia[20]

http://www.usaid.gov/div/portfolio/Zambia-Community-Health-Workers-IPA
DIV contribution: Stage 1 Funding
Award Amount: $99,032
Award Period: 2012 - 2014

Problem: Community health workers (CHWs) are commonly regarded as a potential solution to the shortage of formal health workers throughout sub-Saharan Africa. Recruited from their communities, trained, and then deployed back to their communities, it is thought that CHWs are more likely to have the necessary relationships, local knowledge, and sense of community responsibility to deliver health services to underserved populations in rural areas, where retention of formal health workers is a perennial challenge.

Employing community health workers may help governments address the shortage of health-care providers in Sub-Saharan Africa. However, it is unclear how offering incentives such as career advancement opportunities might affect who applies for the position. For instance, do certain recruitment strategies encourage people with particular personalities or skills to self-select into a job? And if so, how can recruitment strategies be adapted to identify applicants with desirable skills and screen out those who are less likely to perform well on the job?

Innovation: The Government of Zambia worked with researchers to test the effects of two recruitment strategies on applicants' characteristics and job performance. From a sample of 330 CHWs at 165 rural health posts in 48 districts of Zambia, half were randomly chosen to receive recruitment posters emphasizing the "social" benefits of becoming a CHW, such as serving and being a leader in one's community, while the other half received recruitment posters emphasizing "career" benefits, such as opportunities for promotion and professional development.

Impact: Results show that making career incentives, rather than social incentives, salient in recruitment posters attracted workers who were more qualified, performed better on the job, and had similar levels of prosocial preferences. CHWs recruited via career incentives had higher secondary school graduation exam scores and were more likely to quality for university admission. They conducted 30% more patient visits than those recruited through social incentives, and did not achieve these gains by targeting easy-to-reach households or by spending less time on each visit. CHWs in the career incentives group also hosted more than twice as many community meetings as their peers in the social incentives group. The program led to a 31% increase in the number of women giving birth in health centers, a 24% increase in the number of children under five who received health checks,

---

[20] Ashraf, Nava, Oriana Bandiera, and Scott Lee. "Do-Gooders and Go-Getters: Career Incentives, Selection, and Performance in Public Service Delivery." Working Paper, March 2015.

and a 20% increase in the rate of polio immunization. Researchers believe that these effects were driven by high performing CHWs who would not have chosen to apply for the position had they not seen materials emphasizing opportunities for career advancement.

Reach: The Government of Zambia is now exclusively using career incentives to recruit all new community health workers around the country, with the aim of hiring and training 5,000 new CHWs over the next several years. To date, the 500 CHWs recruited using career incentives posters have conducted 49,000 additional household visits over 18 months, reaching approximately 240,100 more people than would have been reached by CHWs in the control group.[21] The Government of Zambia is continuing to work with researchers to identify ways in which this recruitment strategy can be used effectively for other civil service positions.

Likely Future Reach and Sustainability: Prospects for sustainable future reach exist. The Zambian government has already decided to use career incentives to recruit new community health workers around the country. Supported by an $898,892 DIV Stage 2 award, the Government of Zambia is cooperating with researchers to identify other positions where a career incentive-based recruitment strategy can be used for hiring. Zambia's continued use of these incentives for CHW hiring is also likely to remain for the foreseeable future.

Additional Resources:
Academic paper - http://www.povertyactionlab.org/publication/do-gooders-and-go-getters-career-incentives-selection-and-performance-public-service-del
J-PAL Evaluation Summary - http://www.povertyactionlab.org/evaluation/recruiting-and-motivating-community-health-workers-zambia
IGC Video on "Co-Producing Knowledge" - https://www.youtube.com/watch?v=7jtuOIdV3j4

---

[21] Email from Oriana Bandiera 12.11.2014. Total reach calculated using an average household size of 4.9 as reported in the 2007 Demographic and Health Survey: 49,000*4.9 = 240,100.

Envisioning Affordable Eye Care for All

http://www.usaid.gov/div/portfolio/optical-solution-envisioning
DIV contribution: Stage 2 Funding
Award Amount: $585,350
Award Period: 2012 - 2015

Problem: An estimated 544 million people around the world could have their vision restored with a simple pair of reading glasses. For many living in developing countries and suffering from vision loss, a pair of eyeglasses could mean the difference between opportunity and loss of income and quality of life.

Innovation: VisionSpring reaches base of the income pyramid (BoP) customers in rural and peri-urban areas through outreach efforts that provide vision screenings and access to affordable glasses. Its business model supports the sale of glasses to the poorest customers (targeting 70 percent of all customers) with revenue from higher-priced products sold to wealthier customers.

VisionSpring has ten years of experience serving the global BoP optical market including successful implementation of the BoPtical Care Model in El Salvador. DIV supported this program in India, which was designed to reach 1.2 million people in six years.
Each of VisionSpring's 10 "BoPtical Care" Hubs established under this award aimed to reach 12,000 individuals annually with high-quality affordable eye care. With this last-mile distribution system, VisionSpring drove down total costs from $18 to approximately $4 for each pair of glasses, increasing their affordability for BoP customers.

Reach: VisionSpring has sold over 274,000 pairs of glasses to Indian consumers via its BoPtical Care hubs. By the end of 2015, VisionSpring had sold nearly 1 million pairs of glasses in India, and nearly 3 million pairs of glasses through its various distribution streams.

Likely Future Reach and Sustainability: Through its hub distribution system, VisionSpring has provided over 274,000 pairs of glasses to BoP individuals in India, financed by a combination of revenue from customers and donors. It plans to both increase sales and to reduce prices for BoP customers and has increased cost coverage to over 60%. It is likely that reaching millions more people will require a hybrid model that combines revenue from customers and corporations (via CSR projects) with public/donor support.

Providing Electricity to Rural India through Renewable Microgrids

http://www.usaid.gov/div/portfolio/bright-answer-providing
DIV contribution: Stage 2 Funding
Award Amount: $300,000
Award Period: 2011 - 2013

Problem: In India, limited power generation, transmission and distribution infrastructure have made access to electricity a major development challenge. According to estimates from the Government of India, 61 million households, comprising 300 million people, live without power. While the World Resources Institute has estimated the market for off-grid energy products in India at $2 billion, an extremely small fraction of that market is currently being served by NGOs and the private sector. Households off the electrical grid currently rely on low-quality energy sources like kerosene, wood, diesel, candles and disposable batteries. Low cost-effectiveness, detrimental public health effects, and negative environmental impact characterize these sources.

Innovation: Mera Gao Power (MGP) designed a solar-powered, village-level microgrid to provide electricity to off-grid villages in India. Through the microgrid model, renewable power is generated by solar panels and stored in battery banks that charge during the day and are discharged at night. Power is distributed to households through short length distribution lines to high-efficiency LED lights that keep power consumption low. DIV provided Stage 2 funding for MGP to establish its first commercial microgrids. Using DIV funding, MGP was able to improve the cost-effectiveness of its microgrid design, reducing the cost of a microgrid capable of providing power to 50 homes from $3,000 to $1,000. Construction time was reduced from one week to one day per village. A three-person team is currently able to construct the backbone of the microgrid within a few hours and connect customers by the end of the same day.

Reach: Mera Gao has provided electricity to more than 120,000 people.

Likely Future Reach and Sustainability: Mera Gao has raised additional funding. However, it has also faced important challenges, so it is too early to say whether it will prove sustainable.

Developing and Testing a Sustainable Distribution Model for Improved Cookstoves

http://www.usaid.gov/div/portfolio/potential-energy-fueling
DIV contribution: Stage 2 Funding
Award Amount: $1,500,000
Award Period: 2012 - 2015

Problem: It is estimated that each year over 4 million people die from illness related to breathing smoke from cooking fires. Collection of firewood can be time consuming. Many models for high-efficiency stoves exist to replace traditional open fire methods, but few have achieved widespread use or commercial sustainability.

Innovation: Originally the Darfur Stoves Project, Potential Energy was founded as a volunteer organization in 2005 by Dr. Ashok Gadgil, Faculty Senior Scientist and Director of the Environmental Energy Technologies Division of Lawrence Berkeley National Laboratory. Dr. Gadgil led a team of Berkeley scientists and engineers in the development of a Berkeley-Darfur stove, pursuing market-testing and end-user feedback in Darfur. The stove requires half as much firewood as traditional cooking methods.

Using lessons learned from early work on cookstove adoption in Darfur, Potential Energy is pursuing a market creation strategy in Ethiopia. The DIV award was intended to help the organization grow its distribution and marketing network and develop innovative pricing models and flexible financing options for consumers, and help it assess the group's impact and the relative effectiveness of the different marketing strategies it pursues using a randomized control trial, which is ongoing at the time of this writing.

Reach: Since implementation of the DIV grant, Potential Energy has distributed clean cookstoves to nearly 140,000 beneficiaries in Darfur.

Likely Future Reach and Sustainability: The original plan was to obtain revenue from customers. So far, donors have been the main revenue source. It is too early to judge the prospects for future reach.

Table 3

Award Stage	Number of Awards	Total Awarded Value	Fraction Reaching More than 100,000 people	Fraction Reaching More than 1,000,000 people	People Reached[22]	DIV Expenditure per Person Reached
Stage 1 (< $100,000)	23	$2,353,136	17% (4/24)	8% (2/24)	6,723,733	$0.35
Stage 2 (<$1,000,000)	19	$9,557,926	44% (8/18)	11% (2/18)	16,931,044	$0.56
Stage 3 (<$15M)	1	$5,516,606	100% (1/1)	100% (1/1)	1,750,000	$3.15

---

[22] Two innovations (Voter Information Report Cards and CommCare) that reached over 100,000 people received both a Stage 1 and a Stage 2 award. In both of these cases, people reached by those innovations are counted as people reached by Stage 2 awards.